



Projet AP3C
Adaptation des Pratiques Culturelles
au Changement Climatique.

Méthodologie Climatique

Septembre 2020



Avec la participation de



TABLE DES MATIERES

INTRODUCTION

1.	Acquisition des données	p 1
1.1.	Type de données	p 1
1.2.	Réseaux de données	p 1
1.3.	Rapatriement des données	p 2
2.	Reconstitution des données manquantes	p 3
2.1.	Préparation, discussion	p 3
2.2.	Procédures de reconstitution	p 4
3.	Homogénéisation des données	p 5
3.1.	Principes	p 5
3.1.1.	Constitution d'une référence	p 5
3.1.2.	Détection des ruptures	p 5
3.1.3.	Correction des ruptures	p 6
3.1.4.	Processus d'homogénéisation	p 8
3.2.	Détails des procédures	p 8
3.2.1.	Températures	p 8
3.2.1.1.	Homogénéisation de la moyenne	p 8
3.2.1.1.1.	Vague1, réseau sans valeur manquante	p 9
3.2.1.1.2.	Vague2, autres stations	p 15
3.2.1.2.	Homogénéisation de la variance	p 18
3.2.1.1.1.	Vague1, réseau sans valeur manquante	p 20
3.2.1.1.2.	Vague2, autres stations	p 23
3.2.1.3.	Transfert de l'homogénéisation de la variance vers les données quotidiennes	p 25
3.2.2.	ETP (EvapoTranspiration Potentielle)	p 31
3.2.2.1.	Homogénéisation de la moyenne	p 31
3.2.2.1.1.	Vague1, réseau sans valeur manquante	p 31
3.2.2.1.2.	Vague2, autres stations	p 34
3.2.2.2.	Homogénéisation de la variance	p 36
3.2.2.3.	Transfert de l'homogénéisation de la variance vers les données quotidiennes	p 37
3.2.3.	Précipitations	p 41
3.2.3.1.	Homogénéisation de la moyenne	p 42
3.2.3.1.1.	Vague1, réseau sans valeur manquante	p 42
3.2.3.1.2.	Vague2, autres stations	p 45
3.2.3.2.	Homogénéisation de la variance	p 46

3.2.3.3.	Transfert de l'homogénéisation de la variance vers les données quotidiennes	p 47
4.	Analyse organisationnelle des séries de données	p 50
4.1.	Déroulement exploratoire	p 50
4.2.	Détermination des bases de fonctions opérationnelles	p 51
4.2.1.	Organisation de long terme	p 51
4.2.1.1.	Températures	p 51
4.2.1.2.	ETP	p 51
4.2.1.3.	Précipitations	p 51
4.2.2.	Organisation de court terme	p 52
4.2.2.1.	Températures	p 52
4.2.2.2.	ETP	p 52
4.2.2.3.	Précipitations	p 53
4.2.3.	Organisation de la variance	p 54
4.2.3.1.	Températures	p 54
4.2.3.2.	ETP	p 54
4.2.3.3.	Précipitations	p 55
4.3.	Extraction automatisée des organisations	p 55
4.3.1.	Températures	p 55
4.3.2.	ETP	p 58
4.3.3.	Précipitations	p 60
4.3.3.1.	Usage de températures locales	p 61
4.3.3.2.	Usage de températures déportées	p 62
5.	Projections climatiques	p 65
5.1.	Déroulement générique	p 65
5.2.	La question du futur	p 66
5.3.	Températures et ETP	p 66
5.4.	Précipitations	p 73
6.	Les Indicateurs Agro-Climatiques (IAC)	p 81
6.1.	Catégorisation	p 81
6.2.	Exemples de procédures	p 84
6.2.1.	Cumul ou ratio sur période fixe	p 84
6.2.2.	Evénements simples sur période fixe	p 88
6.2.3.	Datation d'un événement, sur période fixe	p 89
6.2.4.	Cumul ou ratio sur période conditionnelle simple (sommés de températures)	p 90
6.2.5.	Conditions complexes sur période conditionnelle	p 91
6.2.6.	Conditionnalités faisant intervenir les statistiques d'un autre indicateur	p 92

7.	Le « Bilan Hydrique Réel » (BHR)	p 95
7.1.	Procédure de calcul	p 95
7.2.	Perspectives	p 98

CONCLUSION

ANNEXES (consulter les fichiers texte, téléchargeables indépendamment)

INTRODUCTION

Le projet AP3C se caractérise par une démarche innovante concernant la méthode utilisée pour l'appréciation du futur climatique et la descente d'échelle infra-régionale. En effet, plutôt que d'utiliser les résultats de la modélisation physique globale du climat, zoomés sur le territoire du Massif central, nous avons conçu et utilisé des projections climatiques locales, de nature statistique, via un générateur stochastique de temps spécifique.

Ce rapport ne revient pas sur les raisons de ce choix et renvoie aux documents concernant l'expertise climatique du projet AP3C¹. De façon très synthétique, rappelons simplement que les projections climatiques habituelles ne sont pas compatibles avec les trajectoires réellement engagées sur le terrain, parce qu'elles ne sont pas conçues pour cela. Il est possible de les utiliser pour une phase de sensibilisation, mais elles sont inappropriées pour une adaptation opérationnelle territoriale aux effets du Changement Climatique, particulièrement durant les premières décennies à venir. Or, c'est bien là tout l'intérêt du projet AP3C : donner aux acteurs locaux et aux agriculteurs toutes les informations nécessaires, exploitation par exploitation et sur un pas de temps resserré, pour une adaptation des pratiques culturelles et des systèmes agricoles du Massif central à moyen terme. C'est un pas en avant vers la nécessaire synergie entre tous les acteurs des filières et des territoires.

Le but de ce document est une description technique détaillée des travaux et procédures à l'œuvre dans AP3C. Ils pourront ainsi être réutilisés, pour projeter le climat et commencer à utiliser ces projections, par les acteurs du Massif central ou sur un autre domaine géographique, pas nécessairement dans un but agricole. La présente publication intervient dans une phase de vulgarisation et d'expérimentations du projet AP3C.

Il est à remarquer qu'un important travail préalable de simplification et d'automatisation des diverses procédures a été conduit pour que la méthode, dans son intégralité, puisse être maîtrisable en quelques semaines par une équipe ayant des compétences basiques en statistiques et en programmation (logiciel R essentiellement).

On décrira d'abord la manière dont les données climatiques quotidiennes ont été acquises et mises en forme, dans une architecture manuelle de base de données.

L'étape suivante, annexe pour une description statique mais cruciale pour une description évolutive, est l'homogénéisation des données. C'est ce qui permet d'assurer la cohérence de l'analyse des organisations chronologiques qui est déroulée à l'issue. Les termes (ou valeurs des paramètres) obtenus dans cette analyse sont ensuite injectés dans des fichiers de configuration de la projection proprement dite.

On détaillera notamment les 2 ou 3 itérations nécessaires à l'obtention de projections « corrigées » car, si la conformité des organisations chronologiques entre projections et observations est intrinsèque à la méthode, il peut se produire des écarts significatifs en termes de distributions de valeurs. Il convient alors de les corriger, d'une manière similaire à ce qui est fait en sortie des modèles climatiques globaux.

A noter que dans tout le cheminement de ce document, on n'abordera que brièvement ce qui relève de la Recherche, c'est-à-dire les différentes étapes exploratoires qui ont conduit à l'établissement et à la stabilisation de la méthode et des procédures. Même si nous y avons souscrit occasionnellement, la démarche académique n'est pas au cœur des préoccupations du projet AP3C.

¹ <https://www.sidam-massifcentral.fr/developpement/ap3c>

« AP3C, un projet qui crée ses propres projections climatiques – mai 2018 »
« Compléments méthodologie climatique AP3C – septembre 2018 »

1. Acquisition des données

Toute activité suivie sur des données commence par un travail de collection ou/et de récupération. Mais, avant de les acquérir, il convient de décider quel type de données sera utilisé et sur quel réseau.

1.1. Type de données

Outre les paramètres climatiques incontournables que sont les températures (T) et la hauteur des précipitations (RR), l'usage agronomique des données nous a orienté vers l'EvapoTranspiration Potentielle (ETP). Ainsi, les problématiques de bilan hydrique peuvent être abordées.

En ce qui concerne le pas de temps, on peut penser à trois types de données :

- Les **données mensuelles** avaient été utilisées au début 2012 à titre de démonstration lors de la préfiguration de l'étude « adaptation des pratiques culturelles au Changement Climatique » menée en Creuse (alias « AP3C-Creuse »). On y gagne une simplicité de mise en œuvre et une grande rapidité d'exécution y compris sur du matériel informatique d'entrée de gamme. Cependant, la limite est de constater que les critères cumulés, sur un mois ou plus, ne permettent pas d'envisager l'analyse de l'évolution d'intempéries pénalisantes. C'est toutefois une échelle qui resterait pertinente pour une approche de sensibilisation aux évolutions moyennes.
- Les **données horaires** restituent très finement le cycle quotidien et permettent d'aborder la quasi-totalité des phénomènes impactants même brefs comme les averses orageuses par exemple. Cependant, par rapport à des valeurs quotidiennes, les observations et projections sont 24 fois plus nombreuses et il devient impératif de séparer l'ETP en ses éléments constitutifs que sont le rayonnement, le vent et l'humidité, en plus de la température. La taille de mémoire vive pour les calculs les plus exigeants devient prohibitive (environ 1To), de même que l'espace de stockage des projections (environ 500To pour l'étude type sur le Massif central).
- Ceci nous a naturellement conduit vers les **données quotidiennes**.

Concernant la période d'étude, elle a été coordonnée avec celle de « AP3C-Creuse », qui démarrait en 1980, mais en y ajoutant 5 années supplémentaires, soit 1980-2015.

Si on devait recommencer une étude de ce type depuis l'origine, il est conseillé de faire une recherche statistique de rupture de pente significative (sur des températures annuelles démarrant vers 1950 par exemple) et de démarrer à cette date, soit plutôt vers 1975 que vers 1980. La date de fin serait, évidemment, la dernière année complète disponible.

1.2. Réseaux de données

Les réseaux de mesures météorologiques utilisés s'étendent non-seulement sur les 11 départements dont les chambres d'agriculture sont partie prenante du projet AP3C mais plus largement sur l'ensemble de l'emprise du territoire « Massif central ». Il ne s'agit pas d'une considération de type administrative par rapport au Commissariat de Massif qui a largement porté ce projet mais principalement d'un intérêt méthodologique concernant l'homogénéisation des données, qui sera présentée au chapitre 3.

En ce qui concerne la densité, il faut séparer la Creuse des autres départements. En effet, le réseau maximaliste étudié dans AP3C-Creuse a été repris à des fins de continuité et de non-dégradation, soit 19 stations T et 34 stations RR. Ailleurs, les limitations de taille se sont appliquées essentiellement pour des questions de coût des données et de quantité de travail à produire, sachant que ce travail serait essentiellement fait de manière expérimentale donc compté en heures de main d'œuvre.

Les critères d'acceptabilité des séries ont été calqués sur ceux de AP3C-Creuse, à savoir une présence effective et continue de données durant une période de 15 ans, verrouillée à 1996-2010, ceci pour que le signal d'évolution tendancielle puisse être au moins détectable sur les observations (en température). Si on devait recommencer ce genre d'étude sans contrainte historique, il est recommandé de garder la période minimale de 15 ans, mais librement positionnée à l'intérieur de la période d'étude totale. Si on souhaite un usage opérationnel des statistiques climatiques produites, il est souhaitable de ne pas retenir les stations fermées depuis plus de 5 ans.

Le réseau complet a été constitué de 32 séries d'ETP, de 158 séries de T (79 de températures minimales Tn et 79 de maximales Tx) et de 124 séries de RR.

1.3. Rapatriement des données

Nous avons choisi de n'utiliser que des séries provenant du portail des données publiques de MétéoFrance (donneespubliques.meteofrance.fr). Cette acquisition se fait via un compte qu'il faut ouvrir en ligne et préalablement abonder avant de l'utiliser. Les références tarifaires se trouvent à cet endroit (https://donneespubliques.meteofrance.fr/?fond=contenu&id_contenu=35), mais à titre indicatif on peut retenir un coût de 0,40€ par lot de 10 données.

Une fois connecté, on se dirige vers le catalogue de climatologie puis vers celui des données climatologiques de base. On demande alors à « en savoir + » sur les données quotidiennes et on clique sur un « achat direct en ligne ». Ensuite il faut mettre les références temporelles, soit du 01011980 au 31122015 dans notre cas. Attention, il convient au préalable d'avoir effectué toutes les vérifications d'ouverture des stations et de présence probable des données grâce aux fiches descriptives des stations. Ceci peut prendre plusieurs heures et, si vous ne le faites pas, vous risquez qu'une partie de votre réseau pressenti ne soit pas exploitable.

Il convient de grouper au maximum les séries commandées pour bénéficier des réductions pour gros volume de données. Je suggère de récupérer un format de fichier texte qui distingue le séparateur de colonne (« ; » par exemple) du séparateur décimal (« , » par exemple). Les fichiers directement issus de ces commandes groupées seront stockés en l'état dans un répertoire qui servira de sauvegarde en cas de mauvaise manipulation ultérieure.

Cependant, comme ces regroupements peuvent mélanger les données ou/et les stations, ils imposeront une désincarcération en fichiers unitaires, paramètre par paramètre et station par station. Il est cependant possible de maintenir les températures minimales et maximales ensemble.

C'est, au plus tard, à ce moment qu'il faut décider d'un regroupement des séries en zones géographiques cohérentes d'un point de vue climatique. Ce regroupement devra avoir un sens aussi bien pour la reconstitution des manques, abordée juste après, que pour l'homogénéisation des données. En ce qui concerne la taille des réseaux, l'expérience acquise dans le domaine de l'homogénéisation incite à privilégier une taille autour de 20 à 30 séries. En dessous de 10 on risque de tomber sur une indécidabilité entre plusieurs cheminements d'homogénéisation et au-delà de 40 la charge mentale devient importante car il y a des choix partiellement manuels à faire entre des représentations graphiques dont le nombre évolue comme le carré du nombre de séries.

Pour ce qui nous concerne, nous avons séparé le Massif central en 3 zones pour les températures et les précipitations. La zone Nord comprend les départements 03, 21, 23, 42, 58, 63, 69, 71 et 89. La zone Centre comprend les départements 07, 15, 30, 43, 48 et 87. Et la zone Sud comprend les départements 11, 12, 19, 34, 46, 81 et 82. Pour l'ETP, la taille limitée du réseau (32 stations) a permis la constitution d'une seule zone globale.

2. Reconstitution des manques

Avant d'appliquer cette reconstitution, il convient de préparer et mettre en forme les données. Ainsi, au sein de chacune des zones géographiques cohérentes, on procède à un regroupement des séries, paramètre par paramètre.

2.1. Préparation, discussion

Notons que la reconstitution proprement dite peut être basique dans la mesure où l'homogénéisation des données, opérée juste après, corrige un certain nombre de défauts des données observées et reconstituées.

Donc, après de nombreux essais, nous avons choisi de procéder à une régression multiple linéaire avec comme prédicteurs potentiels les séries qui ne comportent aucun manque. Ceci permet d'avoir une seule expression de l'estimateur par série à compléter.

Les modalités de la régression varient un peu en fonction du paramètre :

- Pour les températures, aussi bien minimales que maximales, la régression intègre une constante.
- Pour l'ETP, c'est une régression linéaire sans constante, pour assurer que des prédicteurs nuls aboutissent à une estimation nulle également.
- Pour les précipitations, le cas est plus complexe et les recherches ont été les plus longues. Elles ont abouti à une double transformation de variable « faiblement non-réversible ». Cette transformation est de type logarithmique. On part de RR, hauteur des précipitations quotidiennes, et on calcule $LRR = \log_{10}(RR + 0,01 + \text{alea_uniforme}[0;0,2])$ ce qui donne une distribution de forme gaussienne. La régression linéaire est ensuite calculée sans inclusion de constante sur les LRR et, enfin, on procède à la transformation approximativement inverse $RR = \text{Max}(0; 10^{LRR} - 0,11)$. Le terme « 0,11 » s'explique par le fait que, en moyenne, une distribution aléatoire uniforme comprise entre 0 et 0,2 est centrée sur 0,1.

On pourrait considérer que la réversibilité imparfaite est un défaut de la méthode mais c'est le prix à payer pour résoudre, de manière élégante, les difficultés liées à la distinction entre les situations pluvieuses et non-pluvieuses. Habituellement, elle conduit à produire 2 schémas séparés dont il faut gérer la concurrence. Ici on produit, de fait, un continuum statistique de situations avec un seul schéma.

N.B. : Jusque-là, toutes les opérations peuvent se réaliser sous tableur. Mais le nombre de prédicteurs peut être élevé et impliquer une instabilité numérique. C'est la raison pour laquelle nous avons opté dans AP3C pour une sélection progressive ascendante (S.P.A.) de ces prédicteurs. Ce genre de procédé, assez sophistiqué, nécessite l'usage d'un logiciel de statistiques. Originellement, la quasi-totalité des opérations statistiques de AP3C ont été réalisées sous le logiciel commercial SYSTAT®, dans la lignée de ce qui avait été fait pour « AP3C-Creuse ». Pour assurer la portabilité et la reproductibilité du procédé, la très grande majorité des procédures ont été traduites sous R (logiciel libre) et, pour un petit nombre, on s'est assuré que les opérations sont réalisables sous MYSTAT®, une version librement téléchargeable (<https://systatsoftware.com/downloads/download-mystat/>) et utilisable de SYSTAT®, pour un usage non-commercial.

Il est à noter que la S.P.A. automatique laisse encore passer un nombre important de prédicteurs qui apportent très peu d'informations quoi qu'elles soient statistiquement significatives. Nous avons donc été conduits à agir manuellement sur cette S.P.A. en limitant la liste des prédicteurs à ceux qui faisaient très sensiblement diminuer la variance du résidu de régression. Une formulation équivalente en est l'augmentation de la corrélation entre la régression et les valeurs du prédicteur.

Pour donner un ordre de grandeur de la qualité du résultat, on aboutit à une proportion de variance expliquée (soit une corrélation au carré) de 0,92 à 0,985 pour Tx, 0,90 à 0,98 pour Tn, 0,72 à 0,77 pour LRR et 0,90 à 0,98 pour ETP, ceci pour le domaine hors Creuse. Pour la Creuse, où la densité est plus forte, on obtient respectivement 0,97 à 0,995 pour Tx, 0,92 à 0,98 pour Tn, 0,79 à 0,84 pour LRR et 0,97 à 0,98 pour ETP.

2.2. Procédures de reconstitution

On présentera ci-dessous les expressions correspondant au langage BASIC SYSTAT/MYSTAT. On prendra l'exemple de la station de Lurcy-Levis (Code INSEE 03155)

Le repérage des valeurs manquantes peut se faire sous la forme :

LET Manquant=MIS (Tn|Tx|LRR|ETP03155) où MIS est un opérateur booléen de repérage d'un manque (vrai=1 ; faux=0).

On affiche la somme de Manquant, on repère si elle est nulle. Si oui alors la station servira de base à la reconstitution des autres. Si non, on obtient le nombre de valeurs à reconstituer.

La procédure interne de régression multiple linéaire tient compte de la présence ou pas des données.

On imprime ou on stocke les paramètres de la régression. Ensuite on inscrit l'estimation dans la colonne ad-hoc (xxx03155) de manière conditionnelle c'est à dire si la donnée initiale était manquante. Voici les expressions concrètes effectivement utilisées pour chacune des 4 variables climatiques étudiées.

```
IF MIS (Tn03155)=1 THEN LET
Tn03155=0,3867+Tn03060*0,4391+Tn89349*0,2801+Tn230961*0,2660

IF MIS (Tx03155)=1 THEN LET
Tx03155=0,4010+Tx03060*0,5761+Tx89349*0,3091-Tn23142*0,2334+Tx23176*0,3377

IF MIS (LRR03155)=1 THEN LET
LRR03155=-0,1251+LRR03060*0,1536+LRR23025*0,1159+LRR23046*0,0992
+LRR23075*0,1283+LRR23114*0,1281+LRR58062*0,1594
+LRR71472*0,0660+LRR89349*0,1017

IF MIS (ETP03155)=1 THEN LET
ETP03155=ETP03060*0,7672+ETP12145*0,0224-ETP42005*0,0255+ETP870856*0,1787
```

Il reste à sauvegarder le fichier de zone obtenu, avec toutes les variables-colonnes dont les manques ont été comblés au format SYSTAT/MYSTAT standard, puisque ce logiciel sera utilisé également pour l'homogénéisation des données.

On a effectivement appliqué cette méthode de reconstitution à l'ensemble du réseau et on a constitué les fichiers TnTxMNc.syd TnTxMCCc.syd TnTxMSc.syd comprenant les séries de températures complètes des zones Massif Nord, Massif Centre et Massif Sud. Des fichiers équivalents ont été obtenus pour les précipitations (RRMNc.syd RRMCCc.syd RRMSc.syd). Pour l'ETP, le fichier global ETPMfc.syd a été produit.

Il s'agissait jusqu'à présent d'opérations *a priori* non-spécifiques au projet AP3C. Les divers cheminements ont donc été décrits de manière assez lapidaire. Tel n'est plus le cas à partir de l'homogénéisation et c'est la raison pour laquelle on détaillera beaucoup plus les processus dans la suite.

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 (il s'est produit telle chose 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. Ensuite, on recherchera les modifications de comportement des séries par rapport à cette référence et on déterminera leurs datations aussi précisément que possible. Enfin, entre chaque couple de dates successives, il sera nécessaire d'appliquer une correction qui fera disparaître la modification de comportement.

3.1.1. Constitution d'une référence

Avec 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. L'exception serait une modification sciemment organisée, mais on peut raisonnablement espérer qu'elle soit documentée en ce cas. Hormis cette exception, il est clair que la moyenne arithmétique des N stations comporte des ruptures N fois plus petites que celles sur les stations individuelles. Elles peuvent donc être négligées en première approximation. Nous utiliserons donc cette moyenne comme référence.

Cependant, 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 $[1/N]^\circ\text{C}$ au premier ordre, à la même date. Quand on effectuera la détection de rupture, on trouvera un décalage différentiel approximatif 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 (cf. 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%. Imaginons qu'un point du cumul des résidus soit extérieur à l'ellipse de confiance à 99%. Il y a alors moins de 1% de probabilité que la série n'ait pas de rupture c'est à dire que la configuration soit fortuite. Concrètement, une personne non-experte dans la méthode prendra le point qui s'écarte le plus de l'ellipse et 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 des résidus aléatoires) 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. De même, on peut se dispenser des corrections liées à l'organisation des résidus (en particulier l'auto-corrélation). Il faut alors considérer que les ellipses de confiance aléatoires effectivement représentées sont sensiblement plus restrictives (plus proches de l'axe des Y) que la réalité.

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. Ce sont les périodes antérieures qui doivent être corrigées. Pour 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 considéré une correction multiplicative sur les précipitations (pour contraindre une mesure nulle en cas de temps sec) ainsi que sur l'ETP (pour le même genre de raisons) et additive sur les températures en °C (ce qui correspond à un décalage constant d'échelle, classique d'un thermomètre manuel).

Dans la mesure où on s'occupe ici de données quotidiennes, on ne pourra pas se contenter de cette correction sur la moyenne (ou l'espérance). Il faudra également rechercher et corriger d'une manière ou d'une autre, les ruptures de variabilité. Le projet AP3C a approché cette variabilité par l'intermédiaire de la variance en regroupement mensuel des données quotidiennes.

Pour les températures, le processus en 2 étapes a pu être déroulé jusqu'au bout. Après les corrections additives sur les données quotidiennes, les corrections multiplicatives sur la variance mensuelle ont été effectuées et retraduites sur les données quotidiennes.

Pour l'ETP et les précipitations, il y a des distorsions de distribution inacceptables. L'étude de la variabilité consiste simplement en une détection des séries homogènes au sens de la variance et le processus d'homogénéisation. Elle sera complétée dans l'étape ultérieure d'analyse d'organisations chronologiques dont les coefficients seront bornés sur l'intervalle décrit par le réseau des séries homogènes.

Voyons l'expression littérale de ces corrections. Soit T_{stat} la variable météorologique dont les valeurs doivent être homogénéisées, et T_{ref} 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 i ème 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. On peut considérer qu'on représente sur un graphe, le cumul de T_{ref} en X et le cumul de T_{stat} en Y, depuis l'origine des temps, en O, et pour chaque pas de temps. Soit r la pente de la droite reliant le premier et le dernier point de chaque période homogène. On pose :

$$r_{i,d} = \text{Cum}(T_{\text{stat}})_{i,d} / \text{Cum}(T_{\text{ref}})_{i,d} \quad \text{et} \quad r_{f,d} = \text{Cum}(T_{\text{stat}})_{f,d} / \text{Cum}(T_{\text{ref}})_{f,d}$$

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 :

$$\begin{aligned} & (T_{\text{stat}})_{i,a} = (T_{\text{stat}})_{i,d} + k \\ \text{soit} \quad & \text{Cum}(T_{\text{stat}})_{i,a} = \text{Cum}(T_{\text{stat}})_{i,d} + k * \text{nobs}_i \end{aligned}$$

Il vient rapidement

$$\begin{aligned} & (T_{\text{ref}})_{i,a} = (T_{\text{ref}})_{i,d} + k/N \\ \text{soit} \quad & \text{Cum}(T_{\text{ref}})_{i,a} = \text{Cum}(T_{\text{ref}})_{i,d} + k/N * \text{nobs}_i \end{aligned}$$

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

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

d'où finalement

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

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 :

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

Si le coefficient d'homogénéisation est multiplicatif alors :

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

On peut montrer que

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

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

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

d'où finalement

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

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

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

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, on effectue un tracé des cumuls de résidus de l'ensemble des stations par rapport à la référence, qui permet d'isoler la station pour laquelle la rupture principale est la plus nette. On procède ensuite, pour cette station, au calcul du double-cumul ($\text{Cum}(T_{\text{ref}})$ en X, $\text{Cum}(T_{\text{stat}})$ 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.

L'expérience a cependant montré qu'il était nettement préférable de scinder ce travail en deux vagues. Premièrement, on homogénéise station par station celles qui sont sans valeur manquante, avec recalcul systématique de la référence. Deuxièmement, on homogénéise les autres stations, sans changer cette référence. Ainsi, les valeurs manquantes reconstituées assez basiquement ne sont pas incorporées à la référence.

3.2. Détails des procédures

3.2.1. Températures

A titre d'exemple, on choisira les températures maximales et la zone Nord du Massif central. On déroulera l'homogénéisation au sens de la moyenne puis au sens de la variance. L'intégralité des fichiers de commande est disponible, sous forme numérique, dans l'annexe « Fichiers Homogénéisation.ods ».

3.2.1.1. Homogénéisation de la moyenne

Étape 0 : Calcul initial de la référence (*Utilisation du fichier de commande Correct.sync, il se confectionne étape par étape*)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation" (choix d'un répertoire de travail)

REM On s'occupe des Tx.
REM*****

REM*****
REM Etape 0 *
REM*****
(suppression des températures minimales nécessaire si on a mis ensemble les minis et les maxis)
USE "TxMNC.syd"
DROP Tn03060 Tn03155 Tn03180 Tn03248 Tn03285 Tn21584 Tn42005 Tn42019 Tn42170 Tn58062 Tn58145 Tn63003
DROP Tn63345 Tn63354 Tn69107 Tn71482 Tn89349 Tn23001 Tn23008 Tn23013 Tn23021 Tn23025 Tn23030 Tn23031
DROP Tn23045 Tn23067 Tn23075 Tn23079 Tn23093 Tn230961 Tn230967 Tn230968 Tn23106 Tn23118 Tn23142 Tn23176
ESAVE "TxMNH.syd"

USE "TxMNH.syd"
```

```

REM Calcul initial de la référence d'homogénéisation
LET TxREG=AVG(Tx03060,Tx42005,Tx58062,Tx89349,Tx23067,Tx230961,Tx23142,Tx23176)
(il y avait donc 8 stations sans manque dans ce réseau)
RUN
SAVE "TxMNH.syd"
RUN

```

3.2.1.1.1. Vague 1, réseau sans valeur manquante

On rappelle qu'on recalculera la référence à chaque étape, pour ce réseau.

Étape 1 : Recherche et correction de la station comportant la rupture la plus flagrante.

(Utilisation du fichier de commande CumRes1.syc, à confectionner ou adapter avant 1^{er} lancement)

```

FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM La méthode sépare les stations en deux groupes et produit donc l'homogénéisation en deux
REM grandes vagues. La première vague voit le traitement des 8 stations n'ayant aucune valeur
REM manquante sur la période 1980-2015

USE "TxMNH.syd"
REM Station n°1 Vichy-Charmeil (inscription manuelle une fois pour toutes, numérotage à discrétion)
LET X=TxReg (à mettre à jour à chaque étape, mettre TxReg01 après étape 1, TxReg02 après étape 2...)
LET Y=Tx03060 (à mettre à jour après l'étape homogénéisant cette station)
RUN

...

REM tracé avec ajout d'un sous-titre
(TITLE= mettre manuellement la totalité avant la première étape, mise à jour de Hxx à chaque étape.
HHxx signifie xx étapes déjà réalisées.)
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx Vichy-Charmeil H00 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ03Charmeil" /WMF (Choix manuel définitif du nom du fichier graphique)

USE "TxMNH.syd"
REM Station n°7 St Etienne-Bouthéon
LET X=TxReg
LET Y=Tx42005
RUN

...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx StEtienne-Bouthéon H00 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE
,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ42Bouthéon" /WMF

...

USE "TxMNH.syd"
REM Station n°36 La Souterraine
LET X=TxReg
LET Y=Tx23176
RUN

...

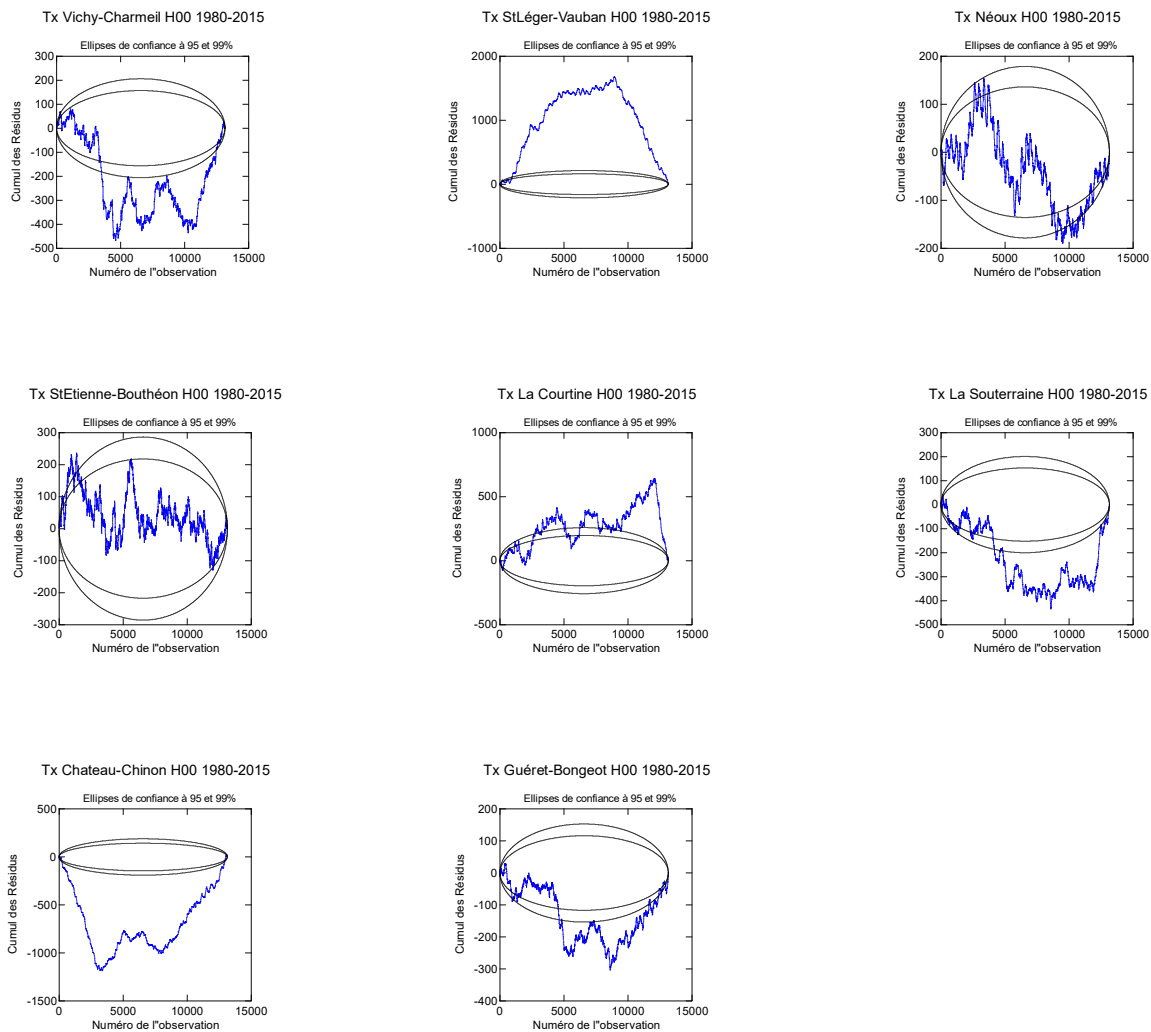
REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx La Souterraine H05 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ23Souterraine" /WMF

```

On consulte l'ensemble des graphiques qui ont été produits (au format vectoriel .wmf) et on repère celui dont les évolutions s'écartent le plus des ellipses de confiance d'homogénéité.

Il n'y a pas eu de formalisation exacte de ce choix car une interversion entre deux stations presque autant in-homogènes est de conséquence négligeable. Par contre, si on corrige en dernière étape de cette première vague une importante rupture alors elle modifiera sensiblement la référence. Ceci risque en retour d'induire une rupture sur toutes les autres stations, c'est-à-dire d'obliger à recommencer le processus.

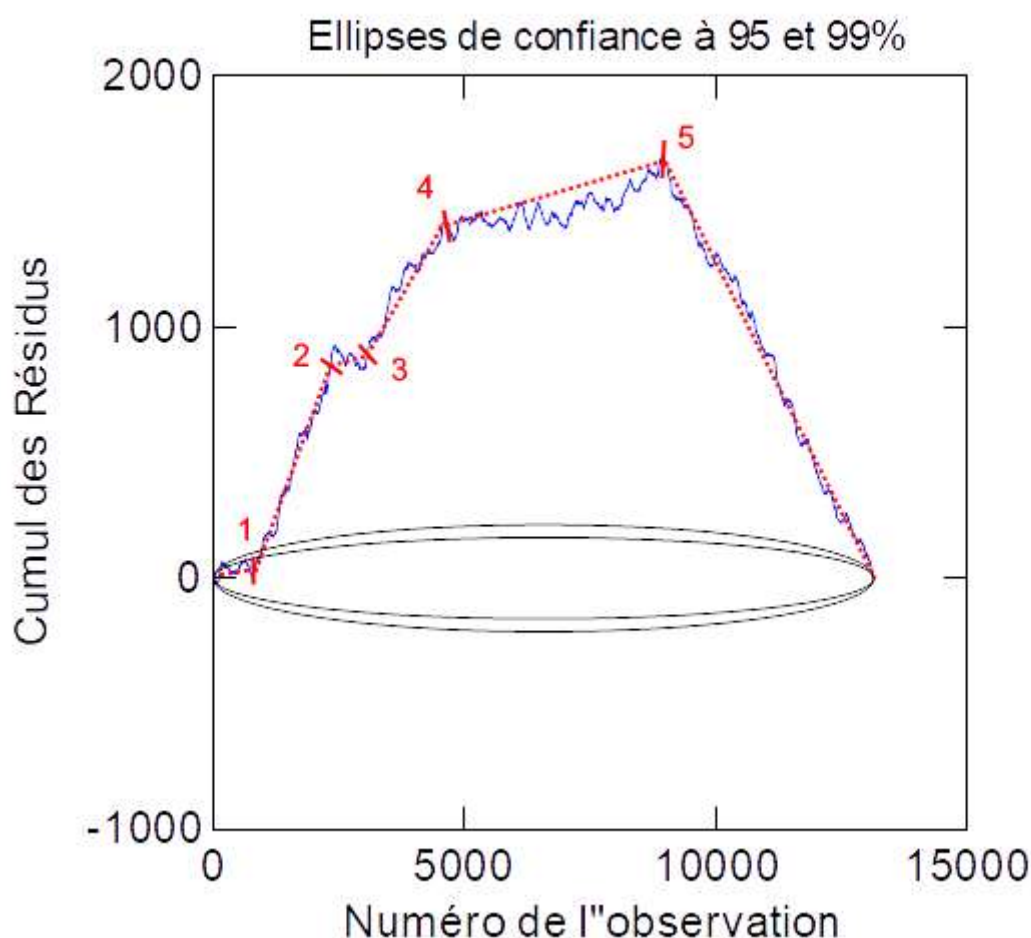
Voici les graphiques pour cette étape :



Le poste de St Léger-Vauban a, sans contestation, l'hétérogénéité la plus nette. Elle peut être traitée de manière fortement différenciée en termes de masses et de temps de travail. Un opérateur qui débute agira rupture par rupture. Un technicien aguerri estimera du premier coup d'oeil le nombre probable de ruptures et les traitera en une seule fois. Nous allons nous placer dans ce dernier cas, sensiblement plus efficace.

Il faut pour cela afficher ou imprimer le graphique correspondant en plein format. On se contraindra impérativement à ré-exécuter la portion de CumulRes1.sys qui est relative à la station choisie, soit St Léger-Vauban ici.

Tx StLéger-Vauban H00 1980-2015



Le travail manuel à effectuer est le repérage approximatif de la position des ruptures. Ce qui revient à tracer mentalement une ligne brisée (ou succession de segments) qui démarre de l'origine et se termine au bout du grand axe des ellipses, en estimant que le cumul des résidus s'en éloignera moins que l'amplitude des ellipses. Ce n'est pas une formulation exacte puisqu'il faudrait plus précisément estimer la réduction de l'amplitude des ellipses de l'étape suivante qui se restreint un peu avec les corrections de ruptures d'homogénéité. Elle est cependant suffisamment précise pour être opérationnelle.

Ici, un opérateur expérimenté repère 5 ruptures potentielles, situées autour des observations numéros 800, 2000, 2700, 4500 et 9000.

Ces approximations nous serviront de guess (première estimation) dans un module de régression qui cherchera la meilleure adéquation, au sens des moindres carrés, d'une ligne brisée à 5 points anguleux à la courbe du cumul des résidus.

(Utilisation du fichier de commande RegresRuptures.syc)

Les numéros d'observations correspondants aux points de ruptures ont été appelés « seuils (S_i) » dans le fichier de commande. Il comprend les commandes pour 1 à 7 ruptures. Il doit être utilisé dans la foulée de CumRes1.syc, sans quitter le logiciel, pour conserver les variables calculées en particulier le cumul des résidus CumRes et le numéro de l'observation NumObs. Rien n'interdit de développer les instructions pour plus de ruptures. Cependant, au plus ce nombre augmente au plus

on risque de prendre une évolution climatique un peu raide pour une rupture. Il est préférable de ne pas utiliser les séries pour lesquelles plus de 7 ruptures significatives apparaîtraient.

```

REM Ajustement du cumul des résidus à l'aide d'une régression linéaire par morceaux
REM*****

REM X:NumObs ; Y:CumRes
REM Démarrage et fin des résidus à zéro soit (X=0,Y=0) et (X=13149 Y=0)
REM Seuils de n ruptures en X=S1,S2...Sn avec S0=0 et Sn+1=13149
REM Equations des n+1 morceaux de droites Y=Ai*X+Bi, morceaux de droites jointives et sécantes en Si-1 et Si
REM
REM Indiquer les valeurs approximatives pour le démarrage de la régression (S=Seuil,A=Pente du segment de droite)

REM Une Rupture
REM Paramètres libres S1,A1 dans l'ordre

NONLIN
SAVE 'RuptParam.syd' / PARAMS
MODEL CumRes = (NumObs<S1)*A1*NumObs,
+ (NumObs>=S1)*A1*S1/(S1-13149)*(NumObs-13149)
ESTIMATE START=10500,-.02
EXIT
USE 'RuptParam.syd'
EXPORT 'RuptParam.txt' /TYPE=ASCII

...

REM Cinq Ruptures
REM Paramètres libres S1,A1,S2,A2,S3,A3,S4,A5,S5,A6 dans l'ordre

NONLIN
SAVE 'RuptParam.syd' / PARAMS
MODEL CumRes = (NumObs<S1)*(A1*NumObs),
+ (NumObs>=S1)*(NumObs<S2)*(A2*NumObs+(A1-A2)*S1),
+ (NumObs>=S2)*(NumObs<S3)*(A3*NumObs+(A2-A3)*S2+(A1-A2)*S1),
+ (NumObs>=S3)*(NumObs<S4)*((A5*(S4-S5)+A6*(S5-13149)-A3*S3-(A2-A3)*S2-(A1-A2)*S1)/(S4-S3)*NumObs,
+A5*(S4-S5)+A6*(S5-13149)-S4*(A5*(S4-S5)+A6*(S5-13149)-A3*S3-(A2-A3)*S2-(A1-A2)*S1)/(S4-S3)),
+ (NumObs>=S4)*(NumObs<S5)*(A5*NumObs+A6*(S5-13149)-A5*S5),
+ (NumObs>=S5)*(A6*(NumObs-13149))
ESTIMATE START=490,0,1380,.8,1900,.2,3500,-.1,6700,-.2
EXIT
USE 'RuptParam.syd'
EXPORT 'RuptParam.txt' /TYPE=ASCII

...

```

Comme on peut le constater, en plus des seuils, il faut aussi avoir une première estimation de la pente des morceaux de droite A_i . Attention, une des valeurs de pente est omise car elle est entièrement déterminée par les autres paramètres. Pour 5 ruptures, c'est la 4ème pente qui est omise (pente du 4ème segment).

Tant que des valeurs stables ne sont pas obtenues, il est préférable de ne pas exécuter les deux dernières lignes car il faudrait recommencer l'exécution de CumRes1.syc pour retrouver les variables nécessaires en entrée de régression par morceaux.

L'exécution est rapide et sans surprise pour une ou deux ruptures. Au-delà, il faut être assez précis dans l'initialisation et de nombreuses difficultés peuvent advenir. Par exemple, deux valeurs très proches (dont les intervalles de confiance se recoupent par exemple) peuvent apparaître. Dans ce cas, il convient de recommencer la régression avec une rupture de moins (attention à la pente omise en initialisation qui ne sera pas nécessairement la même). La convergence peut être en échec. Cela peut être un artefact qui apparaît quand une rupture d'homogénéité se trouve exactement sur une valeur entière stricte de numéro d'observation. En ce cas, on valide manuellement la solution. Parfois, il s'agit d'une convergence lente et il faut augmenter le nombre d'itérations.

Pensez aussi à vérifier que les seuils sont toujours dans un ordre croissant, car la régression par morceaux ne le garantit pas.

Ce processus reste tout de même beaucoup plus rapide qu'une régression entièrement manuelle de l'opérateur, sauf rare exception.

Cette régression semi-automatique produit le fichier de coefficients « RuptParam.txt » qui sera envoyé sur une procédure R.

Il reste à produire les doubles cumuls et à extraire la variable banalisée Y (en fait la température maxi de St-Léger Vauban en ce cas) à corriger.
(Utilisation du fichier de commande DblCumul.syc)

```

FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

...

USE "TxMNH.syd"

LET X=TxReg (à mettre à jour manuellement à chaque étape, TxReg01, TxReg02...)
LET Y=Tx89349 (à mettre à jour après homogénéisation -> Tx89349H01)
LET NumObs=Case
LET Decal0=-40 (décalage du zéro, pour éviter les rebroussements, utile pour une visualisation graphique
des doubles cumuls, qui n'est plus utilisée en mode semi-automatique...)

SAVE "DoubleC1.syd"
RUN
USE "DoubleC1.syd"

BASIC
HOLD
IF Case=1 THEN FOR
    LET CumX=0
    LET CumY=0
NEXT

LET CumX=CumX+X-Decal0
LET CumY=CumY+Y-Decal0

RUN

SAVE "DoubleC2.syd"
RUN
EXIT

MERGE 'DoubleC1.syd' (NUMOBS) 'DoubleC2.syd' (CUMX CUMY)
RUN
EXPORT 'DblCumul.txt' /TYPE=ASCII

MERGE 'DoubleC1.syd' (NUMOBS) 'DoubleC2.syd' (Y)
RUN
EXPORT 'FichY.txt' /TYPE=ASCII

USE "TxMNH.syd"

LET X=TxReg
LET Y=Tx23001
LET NumObs=Case
LET Decal0=-40

...

```

Il faut exécuter uniquement la portion relative à la station considérée (St Léger-Vauban INSEE 89349). Cette procédure fournit les fichiers « DblCumul.txt » et « FichY.txt ».

Les calculs de correction des ruptures d'homogénéité, originellement effectués via tableur et procédures SYSTAT/MYSTAT sont dévolus à une procédure R. Les fichiers « RuptParam.txt », « DblCumul.txt » et « FichY.txt » sont envoyés dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptAdd_ReseauBase.R)

```

# Script de correction de ruptures d'homogénéité, par coefficient additif
#####

# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation, nombre de stations du réseau de référence

Chemin<-"/home/debian/Documents/Rfiles/Homogénéisation/" (A remplir une fois pour toutes)
RefPerFinale<-TRUE (modifiable à chaque étape)
PerRef<-1 (modifiable à chaque étape)
NumEtape<-"Etape01" (respect manuel impératif de la convention typologique, "Etape02", "Etape03"... )
NbStatRef<-8 (nécessaire tant qu'on modifie la référence à chaque étape)

# On importe le fichier des ruptures
NomFichier<-paste(Chemin,"RuptParam.txt",sep="")
RuptParam <-
  read.table(NomFichier,
    header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)

```

```

NbPer<-ncol(RuptParam)/2+1
if (RefPerFinale)
  {PerRef<-ncol(RuptParam)/2+1}

# on importe le fichier des doubles-cumuls
NomFichier<-paste(Chemin,"DblCumul.txt",sep="")
DblCumul <-
  read.table(NomFichier,
    header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)

# on importe le fichier des valeurs à corriger
NomFichier<-paste(Chemin,"FichY.txt",sep="")
Yinit <-
  read.table(NomFichier,
    header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE)

# DESCRIPTION DES CORRECTIONS
#####

# Création du tableau de description des corrections
TabRupt<-matrix(0,nrow=NbPer,ncol=8)
colnames(TabRupt)<-c("Obs","CumX","CumY","DCumX","DCumY","Ri","MoyX","KAdd")

# Remplissage tableau de description des corrections

# Numéro Obs des ruptures
for (i in seq(1,NbPer-1,1))
  {TabRupt[i,1]<-round(RuptParam[1,2*i-1],0)}
TabRupt[NbPer,1]<-nrow(DblCumul)

# Cumuls des X et Y
for (i in seq(1,NbPer,1))
  {LigneCumul<-which(DblCumul[,1]==TabRupt[i,1])
  TabRupt[i,2]<-DblCumul[LigneCumul,2]
  TabRupt[i,3]<-DblCumul[LigneCumul,3]
  }

# Différence des cumuls de X et Y
TabRupt[1,4]<-TabRupt[1,2]
TabRupt[1,5]<-TabRupt[1,3]
for (i in seq(2,NbPer,1))
  {TabRupt[i,4]<-TabRupt[i,2]-TabRupt[i-1,2]
  TabRupt[i,5]<-TabRupt[i,3]-TabRupt[i-1,3]}

# Rapport des différences de cumuls
for (i in seq(1,NbPer,1))
  {TabRupt[i,6]<-TabRupt[i,5]/TabRupt[i,4]}

# Moyenne des X
TabRupt[1,7]<-TabRupt[1,4]/TabRupt[1,1]
for (i in seq(2,NbPer,1))
  {TabRupt[i,7]<-TabRupt[i,4]/(TabRupt[i,1]-TabRupt[i-1,1])}

# Correction additive
for (i in seq(1,NbPer,1))
  {TabRupt[i,8]<-TabRupt[i,7]*(TabRupt[PerRef,6]-TabRupt[i,6])*NbStatRef/(NbStatRef-TabRupt[PerRef,6])}

# Ecriture du fichier de description des corrections
NomFichier<-paste(Chemin,"TabRupt_",NumEtape,".txt",sep="")
write.table(round(TabRupt,6),
  NomFichier, sep=" ",
  col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA", dec=".")

# APPLICATION DES CORRECTIONS
#####

# Création du tableau des valeurs corrigées
YCorr<-matrix(0,nrow=nrow(Yinit),ncol=2)
colnames(YCorr)<-c("NumObs","YCorr")

YCorr[,1]<-Yinit[,1]

# Calcul et application des corrections
for (j in seq(1,TabRupt[1,1],1))
  {YCorr[j,2]<-Yinit[j,2]+TabRupt[1,8]}
for (i in seq(2,NbPer,1))
  {for (j in seq(TabRupt[i-1,1]+1,TabRupt[i,1],1))
  {YCorr[j,2]<-Yinit[j,2]+TabRupt[i,8]}
  }

# Ecriture du fichier des valeurs corrigées
NomFichier<-paste(Chemin,"YCorr.txt",sep="")
write.table(YCorr,
  NomFichier, sep=" ",
  col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA", dec=".")

```

Cette procédure produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt », qui sont à retransmettre dans l’environnement SYSTAT/MYSTAT. Préférez le déplacement à la copie de fichiers ou assurez-vous de supprimer les 2 fichiers dans leur répertoire d’origine après transfert.

Il est conseillé d'imprimer le fichier « TabRupt_Etapexx.txt » pour vérification de conformité et il est impératif de prévoir un répertoire spécifique de sauvegarde. Il contient k+1 lignes où k est le nombre de ruptures. Ces lignes contiennent la position de la rupture (ou de la dernière observation) en colonne 1 et la correction à effectuer sur la période allant du point de rupture précédent (ou de la première observation) au point courant en dernière colonne. Les autres colonnes sont des intermédiaires de calcul qui permettent de vérifier visuellement qu'il n'y a pas d'incohérence. Elles étaient particulièrement utiles dans la version non-automatisée de l'homogénéisation (via l'usage d'un tableur).

Le fichier « YCorr.txt » rejoint le répertoire de travail SYSTAT/MYSTAT pour intégration au fichier de l'ensemble des données à homogénéiser (TxMNH.syd dans notre cas). Attention à bien supprimer les versions antérieures de « YCorr.txt », sinon on pourrait fabriquer un duplicata de la station qui a été homogénéisée précédemment.

(Utilisation du fichier de commande Correct.syc)

```
REM*****
REM Etape 1 *
REM*****
REM Station de St Léger Vauban

IMPORT "YCorr.txt" / TYPE=ASCII
LET Tx89349H01=YCorr (respecter manuellement l'incrément de Hxx)
ESAVE 'YCorr.syd'
MERGE 'TxMNH.syd' 'YCorr.syd' (Tx89349H01)

REM On fait le recalcul du « vecteur régional » (référence)
(à incrémenter et à mettre à jour avec le nom de la nouvelle variable corrigée)
LET TxReg01=AVG(Tx03060,Tx42005,Tx58062,Tx89349H01,Tx23067,Tx230961,Tx23142,Tx23176)

RUN
SAVE "TxMNH.syd"
RUN
```

Étapes 2 à nn : Recherche et correction de la station comportant la rupture la plus flagrante.

(Utilisation du fichier de commande CumRes1.syc)

(Utilisation du fichier de commande RegresRuptures.syc)

(Utilisation du fichier de commande DblCumul.syc)

(Utilisation du fichier de commande CorrecRuptAdd_ReseauBase.R)

(Utilisation du fichier de commande Correct.syc)

Fin de la vague 1 quand les graphes de CumRes1.syc sont tous dans les ellipses de confiance, ou presque, vu qu'elles sont restrictives. Il faut pour cela le vérifier en exécutant CumRes1.syc une dernière fois, de manière isolée. Dans notre cas concret, 5 étapes ont été nécessaires pour corriger 13 ruptures d'homogénéité sur ce réseau.

3.2.1.1.2. Vague 2, autres stations

On rappelle que la référence n'évoluera plus, pour ce réseau. Le processus ressemble beaucoup au précédent mais avec des noms de fichiers de commande parfois distincts.

Étape nn+1 : Recherche et correction de ruptures d'homogénéité, station par station. L'ordre importe peu, respecter de préférence la numérotation croissante si vous en avez adoptée une.

(Utilisation du fichier de commande CumRes2.syc, à confectionner ou adapter avant 1^{er} lancement)

```
REM Cette méthode sépare les stations en deux groupes et produit donc l'homogénéisation en deux
REM grandes vagues. La deuxième vague voit le traitement des 28 stations ayant des valeurs
REM manquantes. Le "vecteur régional" est celui obtenu précédemment et il reste fixe.

REM Deuxième vague d'homogénéisation. Vecteur régional fixe TxReg05.
```

```

USE "TxMNH.syd"
REM Station n°2 Lurcy-Levis (inscription manuelle une fois pour toutes, numérotage à discrétion)
LET X=TxReg05 (on ne met plus à jour la référence)
LET Y=Tx031155 (à mettre à jour après l'étape homogénéisant cette station)
RUN
...

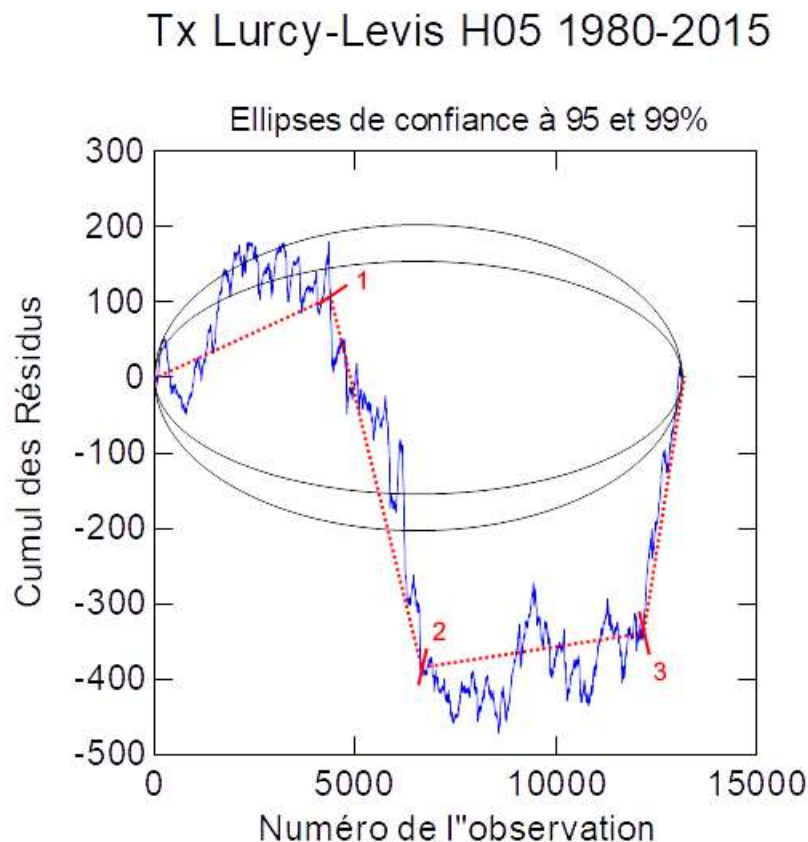
REM tracé avec ajout d'un sous-titre
(TITLE= mettre manuellement la totalité avant la première étape, mise à jour de Hxx à chaque étape.
Hxx signifie xx étapes déjà réalisées.)
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l"observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx Lurcy-Levis H05 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ03LurcyL" /WMF (choix manuel définitif du nom du fichier graphique)
...

USE "TxMNH.syd"
REM Station n°34 Maisonnisses
LET X=TxReg05
LET Y=Tx23118H28
RUN
...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l"observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx Maisonnisses H28 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ23Maisonnisses" /WMF

```

Comme il n'y a plus de choix de stations, il suffit d'afficher seulement le graphe de la station à corriger. Dans notre cas, il s'agit de Lurcy-Levis (INSEE 03155).



Le repérage approximatif des ruptures suggère les numéros d'observation 4000, 6500 et 12000. On les injecte dans le fichier de régression linéaire par morceaux.

(Utilisation du fichier de commande RegresRuptures.syc)

Son utilisation est strictement équivalente à celle du réseau sans manque. On produit le fichier « RuptParam.txt ».

(Utilisation du fichier de commande DblCumul.syc)

Son utilisation est aussi strictement équivalente à celle du réseau sans manque. Attention cependant à réactualiser une dernière fois le nom de la variable de référence, puisque ça n'a pas été fait à l'issue de la dernière étape de la vague 1.

On produit les doubles cumuls et la variable banalisée Y à corriger (la température maxi de Lurcy-Levis en ce cas). On joint les fichiers « DblCumul.txt » et « FichY.txt » au fichier « RuptParam.txt » obtenu juste avant et on les transfère dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptAdd_ReseauSecond.R)

```
# Script de correction de ruptures d'homogénéité, par coefficient additif, réseau secondaire
#####
# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation

Chemin<-" /home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-2
NumEtape<-"Etape06" (bien mettre les numéros d'étape dans la continuité du réseau sans manque)

...
```

L'utilisation en est strictement équivalente à celle du réseau sans manque, sauf qu'on n'a plus à renseigner la taille du réseau.

On produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt ». Pensez à nettoyer le répertoire « R » après transfert vers le répertoire de travail SYSTAT/MYSTAT.

(Utilisation du fichier de commande Correct.syc)

```
REM*****
REM Etape 6 *
REM*****
REM Attention, on ne change plus la référence donc pas de sur-correction ni de recalcul.

REM Station de Lurcy-Levis

IMPORT "YCorr.txt" / TYPE=ASCII
LET Tx03155H06=YCorr (respecter manuellement l'incrément de Hxx)
ESAVE 'YCorr.syd'
MERGE 'TxMNH.syd' 'YCorr.syd' (Tx03155H06)

RUN
SAVE "TxMNH.syd"
RUN
```

Étapes nn+2 à pp : Recherche et correction des ruptures d'homogénéité de la station suivante.

(Utilisation du fichier de commande CumRes2.syc)

On affichera également le cumul des résidus sur la station de l'étape précédente pour vérification de la bonne correction.

(Utilisation du fichier de commande RegresRuptures.syc)

(Utilisation du fichier de commande DblCumul.syc)

(Utilisation du fichier de commande CorrecRuptAdd_ReseauSecond.R)

(Utilisation du fichier de commande Correct.syc)

Fin de la vague 2 à la fin de la liste des stations. On peut relancer une dernière fois CumRes2.sync pour un affichage général final et on peut également mettre à jour DblCumul.sync avec le dernier nom de variable homogénéisée.

Dans notre cas concret, 23 étapes supplémentaires ont été nécessaires pour corriger 49 ruptures d'homogénéité sur ce second réseau. On peut procéder maintenant à l'homogénéisation de leur variance.

3.2.1.2. Homogénéisation de la variance

Deux étapes préalables sont à réaliser :

- Etape 1 : conserver une trace synthétique du travail exécuté sur la moyenne.
- Etape 2 : calculer les variances mensuelles des séries déjà homogénéisées en moyenne.

Pour la première étape, on a choisi de confectionner manuellement un fichier de tableur (ConfigRuptV.ods si c'est le tableur Calc de LibreOffice) avec 2 colonnes comme ceci

Station	NomVarHm
Vichy-Charmeil	TX03060H05
Lurcy-Levis	TX03155H06
Montbeugny	TX03180H07
St-Nicolas	TX03248H08
Tortezais	TX03285H09
Saulieu	TX21584H10
St-Etienne Bout.	TX42005
Boen/Lignon	TX42019
Perreux	TX42170
Chateau-Chinon	TX58062H02
Lormes	TX58145H11
Ambert	TX63003H12
St-Genès	TX63345H13
St-Gervais	TX63354H14
Lamure	TX69107H15
St-Symphorien	TX71482H16
St-Leger Vauban	TX89349H01
Ahun	TX23001H17
Aubusson	TX23008H18
Auzances	TX23013H19
Bénévent	TX23021H20
Bonnat	TX23025H21
Bourganeuf	TX23030
Boussac	TX23031
Chambon/Voueize	TX23045H22
La Courtine	TX23067H03
Dun le Palestel	TX23075H23
Felletin	TX23079H24
Gouzon	TX23093H25
Guéret-Bongéot	TX230961
Guéret-Grancher	TX230967H26
Guéret-Courtille	TX230968
Lépaud	TX23106H27
Maisonnisses	TX23118H28
Néoux	TX23142
La Souterraine	TX23176H04

On remarquera que certaines stations n'ont pas de suffixe Hxx dans leur nom de variable homogénéisée, simplement parce qu'elles étaient déjà homogènes à l'origine. Le nom des stations est libre mais veillez à écrire les **variables avec des majuscules**.

Pour la seconde étape, on fait appel à SYSTAT/MYSTAT.

(Utilisation du fichier de commande VarMens.sys, à confectionner entièrement ou à adapter d'une précédente utilisation)

```
REM*****
REM Programme de fabrication d'écart-type mensuel      *
REM*****

REM Variable Tx

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNH.sysd"

REM Création variable d'indice
let AnMois=Int(Date/100)

REM Transformation des variables pour les rendre plus universelles

let Tx03060Hm=Tx03060H05  (écriture manuelle des noms de variable ou récupération depuis ConfigRuptV.ods)
let Tx03155Hm=Tx03155H06
let Tx03180Hm=Tx03180H07
let Tx03248Hm=Tx03248H08
let Tx03285Hm=Tx03285H09
let Tx21584Hm=Tx21584H10
let Tx42005Hm=Tx42005
let Tx42019Hm=Tx42019
let Tx42170Hm=Tx42170
let Tx58062Hm=Tx58062H02
let Tx58145Hm=Tx58145H11
let Tx63003Hm=Tx63003H12
let Tx63345Hm=Tx63345H13
let Tx63354Hm=Tx63354H14
let Tx69107Hm=Tx69107H15
let Tx71482Hm=Tx71482H16
let Tx89349Hm=Tx89349H01
let Tx23001Hm=Tx23001H17
let Tx23008Hm=Tx23008H18
let Tx23013Hm=Tx23013H19
let Tx23021Hm=Tx23021H20
let Tx23025Hm=Tx23025H21
let Tx23030Hm=Tx23030
let Tx23031Hm=Tx23031
let Tx23045Hm=Tx23045H22
let Tx23067Hm=Tx23067H03
let Tx23075Hm=Tx23075H23
let Tx23079Hm=Tx23079H24
let Tx23093Hm=Tx23093H25
let Tx230961Hm=Tx230961
let Tx230967Hm=Tx230967H26
let Tx230968Hm=Tx230968
let Tx23106Hm=Tx23106H27
let Tx23118Hm=Tx23118H28
let Tx23142Hm=Tx23142
let Tx23176Hm=Tx23176H04

REM Fabrication des écart-types et stockages

BY AnMois

SAVE 'C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNVm.sysd'
(même nom de fichier que données quotidiennes mais avec suffixe vm pour variance mensuelle)
STATS
STATS Tx03060Hm Tx03155Hm Tx03180Hm Tx03248Hm Tx03285Hm Tx21584Hm Tx42005Hm Tx42019Hm,
Tx42170Hm Tx58062Hm Tx58145Hm Tx63003Hm Tx63345Hm Tx63354Hm Tx69107Hm Tx71482Hm,
Tx89349Hm Tx23001Hm Tx23008Hm Tx23013Hm Tx23021Hm Tx23025Hm Tx23030Hm Tx23031Hm,
Tx23045Hm Tx23067Hm Tx23075Hm Tx23079Hm Tx23093Hm Tx230961Hm Tx230967Hm Tx230968Hm,
Tx23106Hm Tx23118Hm Tx23142Hm Tx23176Hm / SD

RUN
```

On peut permuter sans problème les points 1 et 2.

Maintenant, on peut appliquer un programme de travail tout à fait similaire à celui qui a été déroulé pour l'homogénéisation au sens de la moyenne sur les données quotidiennes.

Étape 0 : Calcul initial de la référence (*Utilisation du fichier de commandes CorrectV.syc, il se confectionne étape par étape, comme pour Correct.syc*)

```
REM*****
REM Etape 0 *
REM*****

USE "TxMNvm.syd"

REM Calcul initial de la référence d'homogénéisation
LET TxReg=AVG (Tx03060Hm,Tx42005Hm,Tx58062Hm,Tx89349Hm,Tx23067Hm,Tx230961Hm,Tx23142Hm,Tx23176Hm)

RUN
SAVE "TxMNvm.syd"
RUN
```

3.2.1.2.1. Vague 1, réseau sans valeur manquante

On rappelle qu'on recalculera la référence à chaque étape, pour ce réseau.

Étape 1 : Recherche et correction de la station comportant la rupture la plus flagrante.

(*Utilisation du fichier de commande CumResV1.syc, à confectionner ou adapter avant 1^{er} lancement*)

```
USE "TxMNvm.syd"
REM Station n°1 Vichy-Charmeil
LET X=TxReg (a mettre à jour à chaque étape)
LET Y=Tx03060Hm (le nom de variable initial indique l'homogénéisation de la moyenne, à mettre à jour à chaque étape)
RUN

...

REM tracé avec ajout d'un sous-titre
(remarquer le titre différent pour la variance, mise à jour de Hxx à chaque étape)
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Sigma Mens. Tx Vichy-Charmeil H00 1980-2010' COLOR=2,10,10,10,10
LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "SmTx03Charmeil" /WMF (remarquer le nom différent pour la variance, choix définitif)

...

USE "TxMNvm.syd"
REM Station n°36 La Souterraine
LET X=TxReg
LET Y=Tx23176Hm
RUN

...

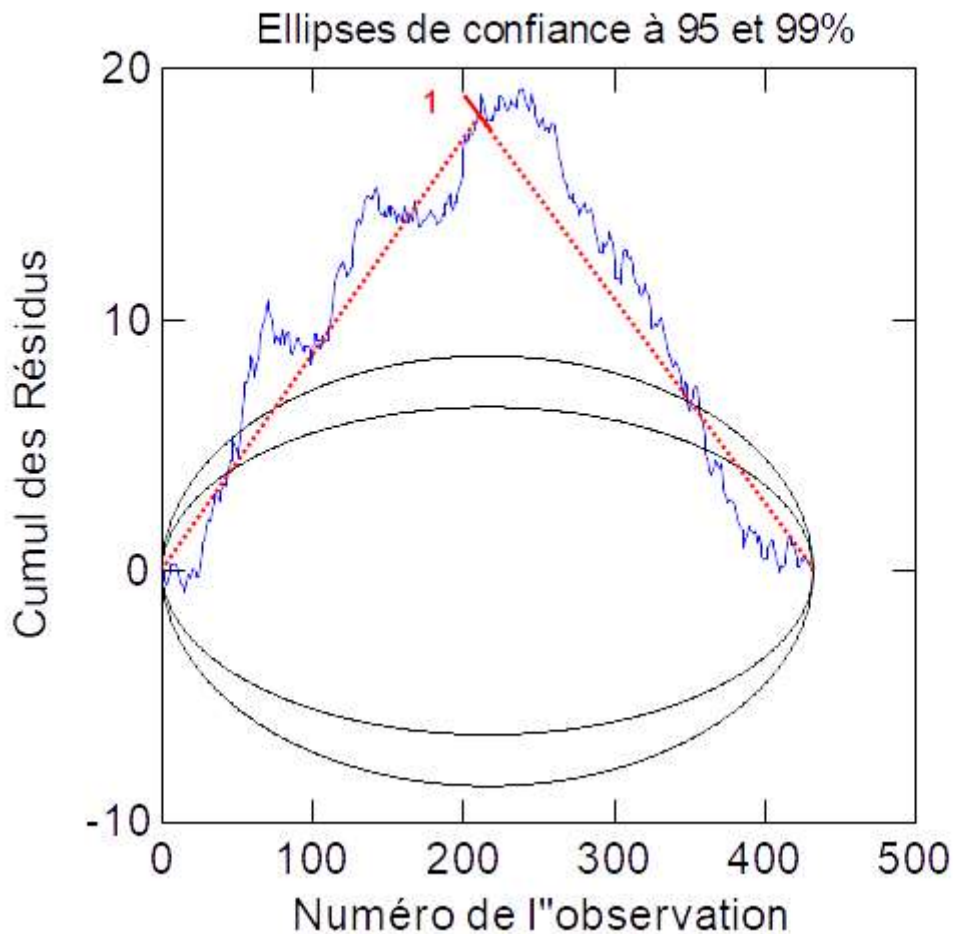
REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Sigma Mens. Tx La Souterraine H00 1980-2010' COLOR=2,10,10,10,10
LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "SmTx23Souterraine" /WMF
```

Etant donné qu'il y a 30 fois moins d'observations, les graphiques sont en général bien plus faciles à interpréter. Ce n'est pas une nécessité de commencer par la même station que pour l'homogénéisation de la moyenne mais dans notre cas concret, c'est à nouveau St Léger-Vauban qui a la plus forte rupture. A noter que ces ruptures de variance ne sont pas nécessairement coordonnées avec les ruptures de moyenne et c'est même, par expérience, un cas très majoritaire. Accessoirement

ceci indique que si vous avez une série homogène au sens de la moyenne, il est peu probable qu'elle soit spontanément homogène au sens de la variance.

Voici le graphe relatif à St Léger-Vauban :

Sigma Mens. StLéger-Vauban H00 1980-2015



Une position de rupture aux alentours de l'observation numéro 205 est évidente. On rappelle que ce 205ème mois correspond approximativement au jour numéro 6250.

Cette valeur est intégrée dans le module de régression linéaire par morceaux, adapté aux variances mensuelles.

(Utilisation du fichier de commande RegresRupturesV.syc)

```
REM Une Rupture
REM Paramètres libres S1,A1 dans l'ordre

NONLIN
SAVE 'RuptParam.syd' / PARAMS
MODEL CumRes = (NumObs<S1)*A1*NumObs,
               + (NumObs>=S1)*A1*S1/(S1-432)*(NumObs-432)
ESTIMATE START=205,0.08
EXIT
USE 'RuptParam.syd'
EXPORT 'RuptParam.txt' /TYPE=ASCII
```

L'exécution en est bien plus rapide que pour les ruptures au sens de la moyenne, car il y a 30 fois moins de données et sensiblement moins de ruptures.

Cette régression semi-automatique produit le fichier de coefficients « RuptParam.txt ».

Il reste à produire les doubles cumulés et à extraire la variable banalisée Y (en fait la variance mensuelle de la température maxi, homogénéisée au sens de la moyenne, de St-Léger Vauban en ce cas) à corriger.

(Utilisation du fichier de commande DblCumulV.syc)

```
USE "TxMNvm.syd"

LET X=TxReg
LET Y=Tx89349Hm
LET NumObs=Case

SAVE "DoubleC1.syd"
RUN
USE "DoubleC1.syd"

BASIC
HOLD
IF Case=1 THEN FOR
  LET CumX=0
  LET CumY=0
NEXT

LET CumX=CumX+X
LET CumY=CumY+Y
RUN

SAVE "DoubleC2.syd"
RUN
EXIT

MERGE 'DoubleC1.syd' (NUMOBS) 'DoubleC2.syd' (CUMX CUMY)
RUN
EXPORT 'DblCumul.txt' /TYPE=ASCII

MERGE 'DoubleC1.syd' (NUMOBS) 'DoubleC2.syd' (Y)
RUN
EXPORT 'FichY.txt' /TYPE=ASCII
```

Ce fichier de commande est utilisé de la même manière que DblCumul.syc. Il produit les mêmes fichiers « DblCumul.txt » et « FichY.txt » qu'on joint à « RuptParam.txt » et qu'on envoie dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptMult_ReseauBase.R)

Notez que le nom a changé car la correction sur la variance est multiplicative.

```
# Script de correction de ruptures d'homogénéité, par coefficient multiplicatif, réseau de base
#####

# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation, nombre de stations du réseau de référence

Chemin<-" /home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-1
NumEtape<-"Etape01"
NbStatRef<-8

...
```

On récupère « TabRupt_Etapexx.txt » et « YCorr.txt », qu'on retransmet dans l'environnement SYSTAT/MYSTAT. Le fichier « YCorr.txt » doit être intégré au fichier de l'ensemble des données à homogénéiser (TxMNvm.syd dans notre cas).

(Utilisation du fichier de commande CorrectV.syc)

```

REM*****
REM Etape 1 *
REM*****
REM Station de Saint-Léger Vauban

IMPORT "YCorr.txt" / TYPE=ASCII
LET Tx89349H01=YCorr (respecter manuellement l'incrément de Hxx)
ESAVE 'YCorr.syd'
MERGE 'TxMNVm.syd' 'YCorr.syd' (Tx89349H01)

REM On fait le recalcul du vecteur régional
(à incrémenter et à mettre à jour avec le nom de la nouvelle variable corrigée)
LET TxReg01=AVG(Tx03060Hm,Tx42005Hm,Tx58062Hm,Tx89349H01,Tx23067Hm,Tx230961Hm,Tx23142Hm,Tx23176Hm)

RUN
SAVE "TxMNVm.syd"
RUN

```

Étapes 2 à nn : Recherche et correction de la station comportant la rupture la plus flagrante.

(Utilisation du fichier de commande CumResV1.syc)
 (Utilisation du fichier de commande RegresRupturesV.syc)
 (Utilisation du fichier de commande DblCumulV.syc)
 (Utilisation du fichier de commande CorrecRuptMult_ReseauBase.R)
 (Utilisation du fichier de commande CorrectV.syc)

Fin de la vague 1 quand les graphes de CumResV1.syc sont tous dans les ellipses de confiance. Il faut pour cela le vérifier en exécutant CumResV1.syc de manière isolée, une dernière fois.

Dans notre cas concret, 3 étapes ont été nécessaires pour corriger 3 ruptures d'homogénéité sur ce réseau.

3.2.1.2.2. Vague 2, autres stations

On rappelle que la référence n'évoluera plus, pour ce réseau. Le processus ressemble beaucoup au précédent mais avec des noms de fichiers de commande parfois distincts.

Étape nn+1 : Recherche et correction de ruptures d'homogénéité, station par station. L'ordre importe peu.

(Utilisation du fichier de commande CumResV2.syc, à confectionner ou adapter avant 1^{er} lancement)

```

REM Deuxième vague d'homogénéisation. Vecteur régional fixe TxReg03.

USE "TxMNVm.syd"
REM Station n°2 Lurcy-Levis
LET X=TxReg03 (a mettre à jour à chaque étape)
LET Y=Tx03155Hm (le nom de variable initial indique l'homogénéisation de la moyenne, à mettre à jour à chaque étape)
RUN

...

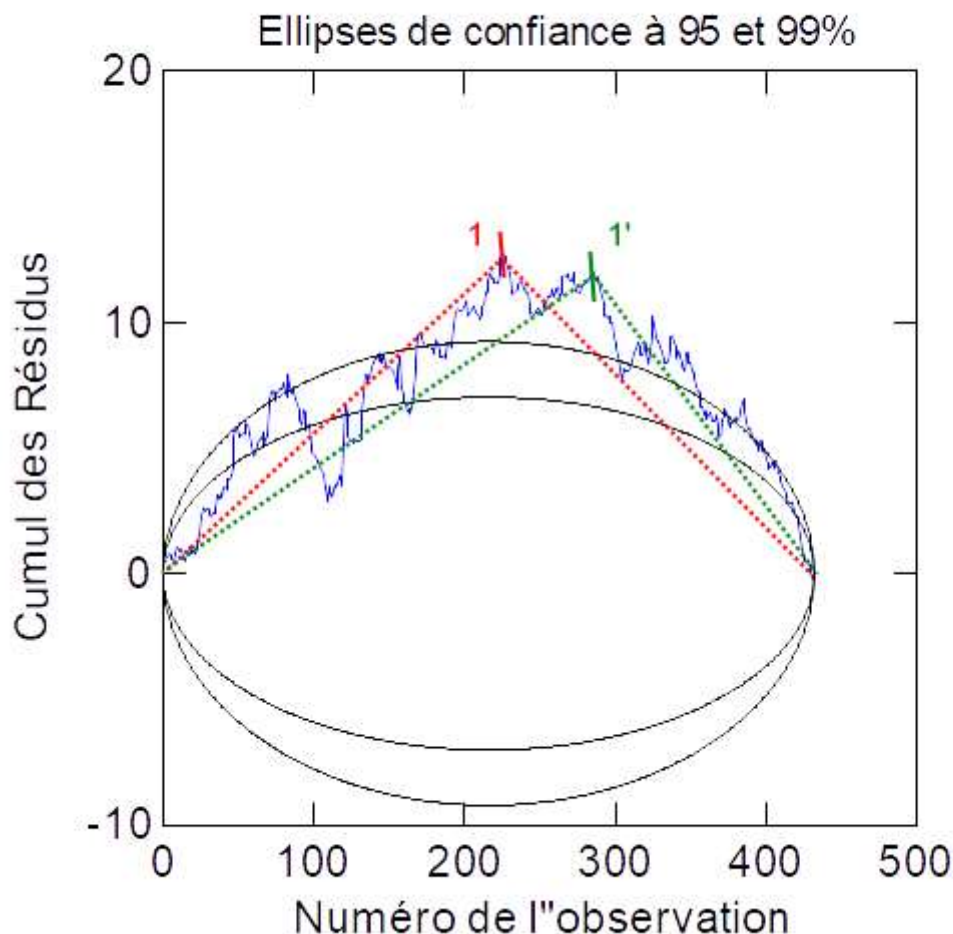
REM tracé avec ajout d'un sous-titre
(remarquer le titre différent pour la variance, mise à jour de Hxx à chaque étape)
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l"observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Sigma Mens. Tx Lurcy-Levis H13 1980-2010' COLOR=2,10,10,10,10
LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "SmTx03LurcyL" /WMF (remarquer le nom différent pour la variance, choix définitif)

...

```

Comme il n'y a plus de choix de stations, il suffit d'afficher seulement le graphe de la station à corriger. Ici, j'ai choisi de représenter la deuxième station de ce réseau alias Montbeugny (INSEE 03180).

Sigma Mens. Tx Montbeugny H03 1980-2015



Le repérage approximatif des ruptures suggère une seule rupture, mais avec 2 possibilités, soit vers l'observation numéro 220, soit vers l'observation numéro 290. On injecte ces deux possibilités dans le fichier de régression linéaire par morceaux. Elles convergent toutes les 2 vers l'observation numéro 268.

(Utilisation du fichier de commande RegresRupturesV.sys)

Son utilisation est strictement équivalente à celle du réseau sans manque. On produit le fichier « RuptParam.txt ».

(Utilisation du fichier de commande DbCumulV.sys)

Son utilisation est aussi strictement équivalente à celle du réseau sans manque. Attention cependant à réactualiser une dernière fois le nom de la variable de référence, puisque ça n'a pas été fait à l'issue de la dernière étape de la vague 1.

On produit les doubles cumuls et la variable banalisée Y à corriger (la variance mensuelle de la température maxi de Montbeugny en ce cas). On joint les fichiers « DbCumul.txt » et « FichY.txt » au fichier « RuptParam.txt » obtenu juste avant et on les transfère dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptMult_ReseauSecond.R)

```
# Script de correction de ruptures d'homogénéité, par coefficient multiplicatif, réseau secondaire
#####
# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation

Chemin<="/home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-2
NumEtape<-"Etape05" (bien mettre les numéros d'étape dans la continuité du réseau sans manque)

« IDEM »
```

L'utilisation en est strictement équivalente à celle du réseau sans manque, sauf qu'on n'a plus à renseigner la taille du réseau.

On produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt »

(Utilisation du fichier de commande CorrectV.syc)

```
REM*****
REM Etape 5 *
REM*****
REM Attention, on de change plus la référence donc pas de recalcul.

REM Station de Montbeugny

IMPORT "YCorr.txt" / TYPE=ASCII
LET Tx03180H05=YCorr
ESAVE 'YCorr.syd'
MERGE 'TxMNVm.syd' 'YCorr.syd' (Tx03180H04)

RUN
SAVE "TxMNVm.syd"
RUN
```

Étapes nn+2 à pp : Recherche et correction des ruptures d'homogénéité de la station suivante.

(Utilisation du fichier de commande CumResV2.syc)

On affichera également le cumul des résidus sur la station de l'étape précédente pour vérification de la bonne correction.

(Utilisation du fichier de commande RegresRupturesV.syc)

(Utilisation du fichier de commande DblCumulV.syc)

(Utilisation du fichier de commande CorrecRuptMult_ReseauSecond.R)

(Utilisation du fichier de commande CorrectV.syc)

Fin de la vague 2 à la fin de la liste des stations. On peut relancer une dernière fois CumResV2.syc pour un affichage général final et on peut également mettre à jour DblCumulV.syc avec le dernier nom de variable homogénéisée.

Dans notre cas concret, 10 étapes supplémentaires ont été nécessaires pour corriger 11 ruptures d'homogénéité sur ce second réseau. Il reste à transférer vers les séries de données quotidiennes les homogénéisations obtenues sur les variances mensuelles.

3.2.1.3. Transfert de l'homogénéisation de la variance vers les données quotidiennes

Le principe général est d'utiliser la formule de modification (normalisation-dénormalisation) de la moyenne et de la variance d'une série X de moyenne m et d'écart-type s vers une série transformée X' de moyenne m' et d'écart-type s'.

Il est patent que la série « (X-m)/s » est normalisée (moyenne 0, écart-type 1). Si on la multiplie par s' et qu'on ajoute m', on fait une dé-normalisation vers X', soit :

$$X' = (X - m) * s' / s + m'$$

Un des intérêts de cette formule simple est d'être applicable de manière instantanée (soit variable avec le temps). Dans le cas qui nous intéresse, il faudra qu'on ait $m' = m$ sur chaque période trouvée homogène en variance car la moyenne ne doit plus être modifiée.

Plusieurs étapes sont nécessaires avant de pouvoir effectuer le transfert de l'homogénéisation de la variance.

Le fichier de configuration des ruptures ConfigRuptV.ods doit être complété manuellement par une 3ème colonne « EtapeHV ». Elle comprend le numéro de l'étape de l'homogénéisation de la variance en deux lettres ou « X » si la série en question était d'origine homogène en variance. Il est impératif que cette 3ème colonne (ainsi que les 2 premières) soit écrite en mode « texte » et pas chiffre.

Pour information, voici ce que nous avons obtenu dans notre exemple :

Station	NomVarHm	EtapeHV
Vichy-Charmeil	TX03060H05	X
Lurcy-Levis	TX03155H06	04
Montbeugny	TX03180H07	05
St-Nicolas	TX03248H08	06
Tortezais	TX03285H09	X
Saulieu	TX21584H10	07
St-Etienne Bout.	TX42005	X
Boen/Lignon	TX42019	X
Perreux	TX42170	X
Chateau-Chinon	TX58062H02	X
Lormes	TX58145H11	X
Ambert	TX63003H12	08
St-Genès	TX63345H13	09
St-Gervais	TX63354H14	10
Lamure	TX69107H15	11
St-Symphorien	TX71482H16	X
St-Leger Vauban	TX89349H01	01
Ahun	TX23001H17	X
Aubusson	TX23008H18	12
Auzances	TX23013H19	13
Bénévent	TX23021H20	X
Bonnat	TX23025H21	X
Bourganeuf	TX23030	X
Boussac	TX23031	X
Chambon/Voueize	TX23045H22	X
La Courtine	TX23067H03	X
Dun le Palestel	TX23075H23	X
Felletin	TX23079H24	X
Gouzou	TX23093H25	X
Guéret-Bongeot	TX230961	X
Guéret-Grancher	TX230967H26	X
Guéret-Courtille	TX230968	X
Lépaud	TX23106H27	X
Maisonnisses	TX23118H28	X
Néoux	TX23142	03
La Souterraine	TX23176H04	02

Exporter ensuite ce fichier au format .csv, en exigeant que chaque texte soit entouré de guillemets. Pour information, voici ce que les premières lignes de « ConfigRuptV.csv » doivent donner :

```
"Station","NomVarHm","EtapeHV"  
"Vichy-Charmeil","TX03060H05","X"  
"Lurcy-Levis","TX03155H06","04"  
...
```

Exporter également au format .txt le fichier SYSTAT/MYSTAT de la fin d'homogénéisation de la moyenne, soit « TxMNH.syd » vers « TxMNH.txt » dans notre exemple.

Ensuite, on transférera les 2 fichiers « ConfigRuptV.csv » et « TxMNH.txt » sous l'environnement R, avec tous les fichiers « TabRupt_Etapexx.txt » relatif à l'homogénéisation de la variance. Attention, il est impératif qu'il y ait au moins un fichier TabRupt_Etape01.txt. Ceci ne se produit pas spontanément si toutes les séries étaient déjà homogènes en variance à l'origine ou si on a décidé de ne pas homogénéiser en variance (voir les variables autres que températures). Dans ce dernier cas, le fichier TabRupt_Etape01.txt ne sert qu'à y lire le nombre de données mensuelles. On devra donc en créer un minimaliste avec "Obs" en ligne 1 et le nombre de mois en ligne 2.

(Utilisation du fichier de commande AppliCorVarDonnéesQ.R)

Cette procédure fournit en sortie le fichier « ListRuptV.txt » qui contient l'ensemble des périodes homogènes au sens de la variance et les corrections appliquées. Il sert essentiellement à faire des vérifications de conformité visuelles, dans cette version semi-automatisée. La procédure fournit surtout le fichier des séries Q homogénéisées au sens de la variance (et de la moyenne) soit « TxMNHV.txt » dans notre exemple.

Ces 2 fichiers sont ensuite transférés sous l'environnement SYSTAT/MYSTAT où le fichier des séries homogénéisées est converti au format du logiciel de statistiques soit « TxMNHV.syd » dans notre exemple.

Comme l'ensemble de ces manipulations est assez complexe et, encore en partie, effectué manuellement, des vérifications de bon traitement sont effectuées.

On vérifie d'abord que les séries homogénéisées finales sont toujours homogènes au sens de la moyenne. Pour cela, on relance un processus d'homogénéisation de la moyenne.

(Utilisation du fichier de commande CumResTest.syc)

Ce fichier traite de l'ensemble du réseau (avec et sans manque au départ).

```
REM Paramètre Tx  
  
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"  
  
REM Création de la référence spécifique à cette méthode par sous-réseau sans manque  
USE "TxMNHV.syd"  
LET TxRegHmv=AVG(Tx03060Hmv, Tx42005Hmv, Tx58062Hmv, Tx89349Hmv, Tx23067Hmv, Tx230961Hmv, Tx23142Hmv, Tx23176Hmv)  
SAVE "TxMNHV.syd"  
RUN  
  
USE "TxMNHV.syd"  
REM Station n°1 Vichy-Charmeil  
LET X=TxRegHmv  
LET Y=Tx03060Hmv  
RUN  
  
...
```



```

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l"observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx Vichy-Charmeil Hmv 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ03Charmeil" /WMF

...

USE "TxMNHV.syd"
REM Station n°36 La Souterraine
LET X=TxRegHmv
LET Y=Tx23176Hmv
RUN

...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l"observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='Tx La Souterraine Hmv 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "TxQ23Souterraine" /WMF

```

On affiche l'ensemble des graphiques et on vérifie qu'ils ont peu évolué par rapport à la fin du processus d'homogénéisation de la moyenne. S'il y a évolution sensible et qu'une nouvelle rupture est apparue, c'est très probablement dû à une erreur de manipulation.

Si le test est probant, on relance l'homogénéisation de la variance.

(Utilisation du fichier de commande VarMens.syc)

Attention on n'exécute que la dernière partie , relative aux vérifications.

```

REM Pour vérification de la non-altération de l'homogénéisation lors du processus de correction.
USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNHV.syd"

REM Création variable d"indice
LET AnMois=Int(Date/100)

BY AnMois

SAVE 'C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNvmTest.syd'
STATS
STATS Tx03060Hmv Tx03155Hmv Tx03180Hmv Tx03248Hmv Tx03285Hmv Tx21584Hmv Tx42005Hmv Tx42019Hmv,
      Tx42170Hmv Tx58062Hmv Tx58145Hmv Tx63003Hmv Tx63345Hmv Tx63354Hmv Tx69107Hmv Tx71482Hmv,
      Tx89349Hmv Tx23001Hmv Tx23008Hmv Tx23013Hmv Tx23021Hmv Tx23025Hmv Tx23030Hmv Tx23031Hmv,
      Tx23045Hmv Tx23067Hmv Tx23075Hmv Tx23079Hmv Tx23093Hmv Tx230961Hmv Tx230967Hmv Tx230968Hmv,
      Tx23106Hmv Tx23118Hmv Tx23142Hmv Tx23176Hmv / SD

RUN

```

(Utilisation du fichier de commande CumResVTest.syc)

Ce fichier traite de l'ensemble du réseau (avec et sans manque au départ).

```

REM Paramètre Tx
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM Création de la référence spécifique à cette méthode par sous-réseau sans manque
USE "TxMNvmTest.syd"
LET TxRegHmv=AVG(Tx03060Hmv, Tx42005Hmv, Tx58062Hmv, Tx89349Hmv, Tx23067Hmv, Tx230961Hmv, Tx23142Hmv, Tx23176Hmv)
SAVE "TxMNvmTest.syd"
RUN

USE "TxMNvmTest.syd"
REM Station n°1 Vichy-Charmeil
LET X=TxRegHmv
LET Y=Tx03060Hmv
RUN

```

```

...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l"observation',

```

```

XMIN=0 YLABEL='Cumul des Résidus' TITLE='Sigma Mens. Tx Vichy-Charmeil Hmv 1980-2015' COLOR=2,10,10,10,10
LEGEND=NONE ,
SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "SmTx03Charmeil" /WMF

```

On affiche l'ensemble des graphiques et on vérifie qu'ils n'ont pas évolué par rapport à la fin du processus d'homogénéisation. S'il y a évolution, c'est très probablement dû à une erreur de manipulation.

En plus de ces vérifications individuelles, on peut également effectuer des vérifications collectives, qui cherchent à s'assurer d'une cohérence spatiale de certaines caractéristiques. Dans la mesure où nous étudions le changement climatique sur une période où il est censé être déjà sensible, une vérification basique de cohérence de tendance (en régression linéaire simple) a été produite.

(Utilisation du fichier de commande VerifTendHom.syc)

Les différents numéros INSEE des stations sont à inscrire manuellement.

```

REM Calculs de regressions linéaires pour contrôle de l'homogénéisation
REM*****

```

```

FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

```

```

REM Evolution tendancielle annuelle, toutes stations
REM*****
REM Données quotidiennes **

```

```

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNH.syd"

```

```

LET Siecle=case/36524

```

```

REM Données Q non-homogénéisées

```

```

REGRESS
MODEL Tx03060 = CONSTANT+SIECLE
ESTIMATE
...

```

```

MODEL Tx23176 = CONSTANT+SIECLE
ESTIMATE
EXIT

```

```

REM Données Q homogénéisées moyenne

```

```

REGRESS
MODEL Tx03060H05 = CONSTANT+SIECLE
ESTIMATE
...

```

```

MODEL Tx23176H04 = CONSTANT+SIECLE
ESTIMATE
EXIT

```

```

REM Données Q homogénéisées variance

```

```

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNHV.syd"

```

```

LET Siecle=case/36524

```

```

REGRESS
MODEL Tx03060Hmv = CONSTANT+SIECLE
ESTIMATE
...

```

```

MODEL Tx23176Hmv = CONSTANT+SIECLE
ESTIMATE
EXIT

```

```

REM Données mensuelles variance**

```

```

REM Données homogénéisées moyenne

```

```

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNVm.syd"
LET Siecle=case/1200

```

```

REGRESS
MODEL Tx03060Hm = CONSTANT+SIECLE
ESTIMATE
...

```

```

MODEL Tx23176Hm = CONSTANT+SIECLE
ESTIMATE
EXIT

REM Données homogénéisées variance

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNvmTest.syd"
LET Siecle=case/1200

REGRESS
MODEL Tx03060Hmv = CONSTANT+SIECLE
ESTIMATE
...

MODEL Tx23176Hmv = CONSTANT+SIECLE
ESTIMATE
EXIT

REM Evolution tendancielle saisonnière, moyenne zonale
REM*****

REM Données Q homogénéisées variance

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNHV.syd"
LET Siecle=case/36524
LET AnMois=int(Date/100)
LET Mois=AnMois-100*int(AnMois/100)

REM Rappel annuel

REGRESS
MODEL TxZonHmv = CONSTANT+SIECLE
ESTIMATE
EXIT

REM Saison par saison

SELECT Mois>=3 AND Mois<=5
REGRESS
MODEL TxZonHmv = CONSTANT+SIECLE
ESTIMATE
EXIT
...

SELECT Mois>=12 OR Mois<=2
REGRESS
MODEL TxZonHmv = CONSTANT+SIECLE
ESTIMATE
EXIT

REM Données mensuelles variance homogénéisées

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\TxMNvmTest.syd"
LET TxZonHmv=AVG(Tx03060Hmv..Tx23176Hmv)
LET Siecle=case/1200
LET Mois=AnMois-100*int(AnMois/100)

REM Rappel annuel

REGRESS
MODEL TxZonHmv = CONSTANT+SIECLE
ESTIMATE
EXIT

REM Saison par saison

SELECT Mois>=3 AND Mois<=5
REGRESS
MODEL TxZonHmv = CONSTANT+SIECLE
ESTIMATE
EXIT
...

SELECT Mois>=12 OR Mois<=2
REGRESS
MODEL TxZonHmv = CONSTANT+SIECLE
ESTIMATE
EXIT

```

Une fois les homogénéisations effectuées, vérifier que les divergences de tendance inter-postes sont beaucoup plus restreintes.

On peut alors dérouler, de manière plus succincte, les procédures d'homogénéisation de données relatives aux autres paramètres.

3.2.2. ETP (EvapoTranspiration Potentielle)

On choisira les exemples dans ce qui a été fait pour l'ensemble de la zone Massif central, car l'homogénéisation de l'ETP s'y est faite globalement. On enchaîne l'homogénéisation au sens de la moyenne puis au sens de la variance. La plupart des opérations étant similaires à celles concernant les températures, les explications seront plus succinctes.

3.2.2.1. Homogénéisation de la moyenne

Étape 0 : Calcul initial de la référence (*Utilisation du fichier de commande Correct.sys, il se confectionne étape par étape*)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM On s'occupe des ETP
REM*****

REM*****
REM Etape 0 *
REM*****

USE "ETPMfc.sys" (c'est le fichier dont les valeurs manquantes ont été reconstituées)

REM Calcul initial de la référence d'homogénéisation
LET ETPReg=AVG(ETP03060,ETP12145,ETP42005,ETP46127,ETP870856)
(il y avait donc 5 stations sans manque dans ce réseau)
RUN
SAVE "ETPMfH.sys"
RUN
```

3.2.2.1.1. Vague 1, réseau sans valeur manquante

On rappelle qu'on recalculera la référence à chaque étape, pour ce réseau.

Étape 1 : Recherche et correction de la station comportant la rupture la plus flagrante.

(*Utilisation du fichier de commande CumRes1.sys, à confectionner ou adapter avant 1^{er} lancement*)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM La méthode sépare les stations en deux groupes et produit donc l'homogénéisation en deux
REM grandes vagues. La première vague voit le traitement des 5 stations n'ayant aucune valeur
REM manquante sur la période 1980-2015

USE "ETPMfH.sys"
REM Station n°1 Vichy-Charmeil
LET X=ETPReg
LET Y=ETP03060
RUN

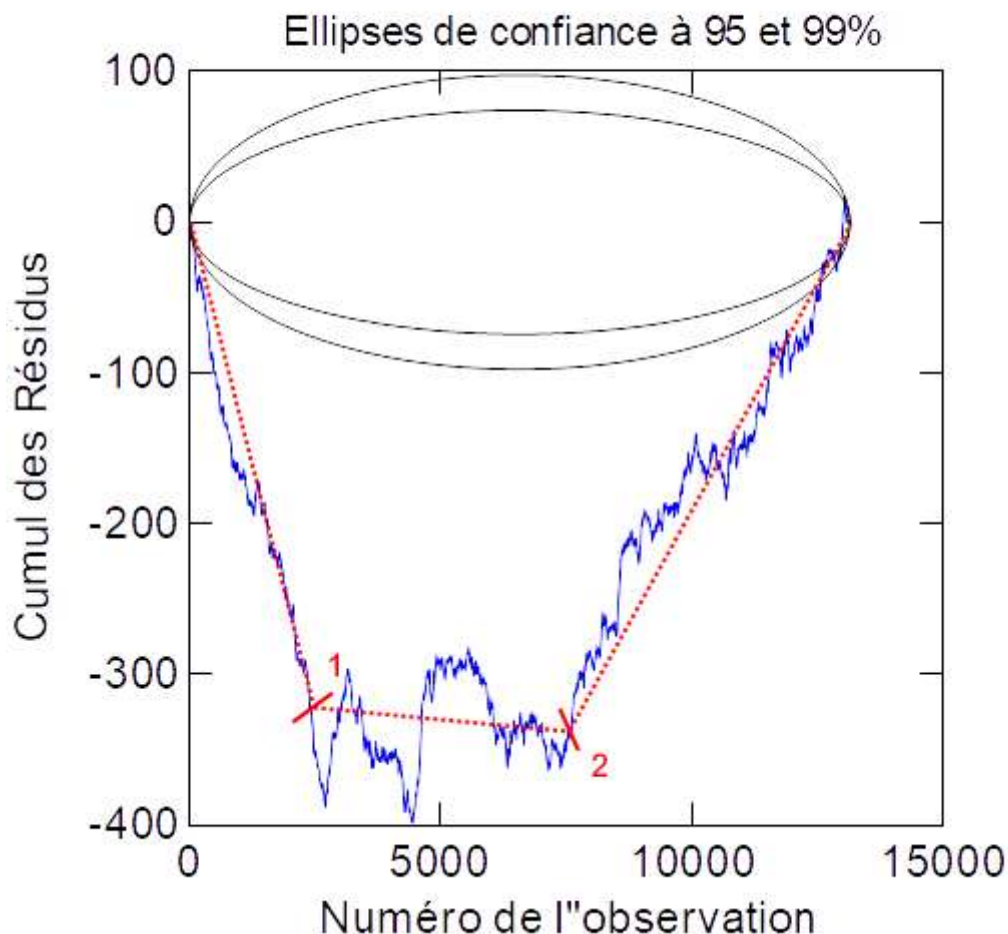
...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='ETP Vichy-Charmeil H00 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "ETPQ03VichyC" /WMF
...
```

On consulte l'ensemble des graphiques qui ont été produits par ce fichier de commande et on repère celui dont les évolutions s'écartent le plus des ellipses de confiance d'homogénéité.

La station de St Etienne-Bouthéon est la plus hétérogène. On se contraint à ré-exécuter la portion de CumulRes1.syc qui est relative à cette station et on affiche ou imprime le graphe en plein format. Le voici :

ETP St Etienne Bouthéon H00 1980-2015



C'est le même travail que pour les températures, on repère approximativement les ruptures, ici 2 ruptures en positions 2500 et 7500.

(Utilisation du fichier de commande RegresRuptures.syc)

Identique aux températures. On produit le fichier de coefficients « RuptParam.txt ».

(Utilisation du fichier de commande DoubleCumul.syc)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"
```

```
...
```

```
USE "ETPMfH.syd"
```

```
LET X=ETPReg
```

```
LET Y=ETP42005
```

```
LET NumObs=Case
LET Decal0=0 (remarque l'absence d'usage du décalage du zéro car l'ETP n'est jamais négative)
...
```

Il faut exécuter uniquement la portion relative à la station considérée (St Etienne-Bouthéon INSEE 42005). Cette procédure fournit les fichiers « DblCumul.txt » et « FichY.txt ».

Les fichiers « RuptParam.txt », « DblCumul.txt » et « FichY.txt » sont envoyés dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptMult_ReseauBase.R)

Veillez noter qu'on utilise pour l'ETP une correction multiplicative dès l'homogénéisation de la moyenne. Ce choix n'est pas spontané, il résulte d'une réflexion sur la nature de la variable et aussi d'une comparaison avec une tentative d'homogénéisation additive sur une variable transformée en logarithme.

Dans ce dernier cas, tout semble se dérouler correctement, jusqu'aux vérifications de cohérences géographiques de tendance, sur la variable transformée. Mais ces mêmes vérifications aboutissent à des tendances violemment divergentes, quand on repasse à la variable ETP d'origine « dé-transformée ».

Néanmoins, cette tentative de transformation de variable n'a pas été inutile dans le sens où elle a été reprise, sous une forme plus élaborée, pour les précipitations (voir plus loin).

```
# Script de correction de ruptures d'homogénéité, par coefficient multiplicatif, réseau de base
#####
# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation, nombre de stations du réseau de référence

Chemin<-" /home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-1
NumEtape<-"Etape01"
NbStatRef<-5

...
```

Cette procédure produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt », qui sont à retransmettre dans l'environnement SYSTAT/MYSTAT. Le fichier « TabRupt_Etapexx.txt » est à contrôler et à stocker dans un répertoire spécifique. Le fichier « YCorr.txt » rejoint le répertoire de travail pour intégration au fichier de l'ensemble des données à homogénéiser (ETPMfH.syd dans notre cas).

(Utilisation du fichier de commande Correct.syc)

```
REM*****
REM Etape 1 *
REM*****
REM Station de St Etienne Bouthéon

IMPORT "YCorr.txt" / TYPE=ASCII
LET ETP42005H01=YCorr
ESAVE 'YCorr.syd'
MERGE 'ETPMfH.syd' 'YCorr.syd' (ETP42005H01)

REM On fait le recalcul du vecteur régional
LET ETPReg01=AVG(ETP03060,ETP12145,ETP42005H01,ETP46127,ETP870856)

RUN
SAVE "ETPMfH.syd"
RUN
```

Étapes 2 à nn : Recherche et correction de la station comportant la rupture la plus flagrante.

(Utilisation du fichier de commande CumRes1.syc)

(Utilisation du fichier de commande RegresRuptures.syc)

(Utilisation du fichier de commande *DbiCumul.syc*)
(Utilisation du fichier de commande *CorrecRuptMult_ReseauBase.R*)
(Utilisation du fichier de commande *Correct.syc*)

Fin de la vague 1 quand les graphes de *CumRes1.syc* sont tous dans les ellipses de confiance. Il faut pour cela le vérifier en exécutant *CumRes1.syc* de manière isolée, une dernière fois.

Dans notre cas concret, 5 étapes ont été nécessaires pour corriger 16 ruptures d'homogénéité sur ce réseau.

3.2.2.1.2. Vague 2, autres stations

On rappelle que la référence n'évoluera plus, pour ce réseau. Le processus ressemble beaucoup au précédent mais avec des noms de fichiers de commande parfois distincts.

Étape nn+1 : Recherche et correction de ruptures d'homogénéité, station par station. L'ordre importe peu.

(Utilisation du fichier de commande *CumRes2.syc*, à confectionner ou adapter avant 1^{er} lancement)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM Cette méthode sépare les stations en deux groupes et produit donc l'homogénéisation en deux
REM grandes vagues. La deuxième vague voit le traitement des 27 stations ayant des valeurs
REM manquantes. Le "vecteur régional" est celui obtenu précédemment et il reste fixe.

REM Deuxième vague d'homogénéisation. Vecteur régional fixe ETPReg05.

USE "ETPMfH.syd"
REM Station n°2 Lurcy-Levis
LET X=ETPReg05
LET Y=ETP03155
RUN

...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='ETP Lurcy-Levis H05 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "ETPQ03LurcyL" /WMF

...
```

On affiche uniquement le graphe de la station considérée. Ici, il s'agit de Lurcy-Levis (INSEE 03155), pour laquelle on repère des ruptures probables aux positions approximatives 4000, 4600, 5200, 7500, 10500 et 12000. Cette pléthore de ruptures n'est pas à considérer de prime abord. Il faut d'abord essayer un nombre potentiellement plus faible et augmenter progressivement en cas d'échec (courbe de cumul des résidus toujours sensiblement en dehors des ellipses de confiance). Avec un nombre de ruptures aussi important, la stabilisation du processus de régression est très sensible et peut prendre un temps (de calcul et d'expertise) conséquent.

On produit le fichier « RuptParam.txt ».

(Utilisation du fichier de commande *DbiCumul.syc*)

Son utilisation est aussi strictement équivalente à celle du réseau sans manque. Attention cependant à réactualiser une dernière fois le nom de la variable de référence, puisque ça n'a pas été fait à l'issue de la dernière étape de la vague 1.

On produit les doubles cumuls et la variable banalisée Y à corriger (l'ETP de Lurcy-Levis en ce cas). On joint les fichiers « DblCumul.txt » et « FichY.txt » au fichier « RuptParam.txt » obtenu juste avant et on les transfère dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptMult_ReseauSecond.R)

```
# Script de correction de ruptures d'homogénéité, par coefficient multiplicatif, réseau secondaire
#####

# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation

Chemin<-"~/home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-2
NumEtape<-"Etape06" (bien mettre les numéros d'étape dans la continuité du réseau sans manque)

« IDEM »
```

L'utilisation en est strictement équivalente à celle du réseau sans manque, sauf qu'on n'a plus à renseigner la taille du réseau.

On produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt »

(Utilisation du fichier de commande Correct.syc)

```
REM*****
REM Etape 6 *
REM*****
REM Station de Lurcy-Levis

REM Attention, on ne change plus la référence donc pas de sur-correction ni de recalcul.

IMPORT "YCorr.txt" / TYPE=ASCII
LET ETP03155H06=YCorr (respecter manuellement l'incrément de Hxx)
ESAVE 'YCorr.syd'
MERGE 'ETPMfH.syd' 'YCorr.syd' (ETP03155H06)

RUN
SAVE "ETPMfH.syd"
RUN
```

Étapes nn+2 à pp : Recherche et correction des ruptures d'homogénéité de la station suivante.

(Utilisation du fichier de commande CumRes2.syc)

On affichera également le cumul des résidus sur la station de l'étape précédente pour vérification de la bonne correction.

(Utilisation du fichier de commande RegresRuptures.syc)

(Utilisation du fichier de commande DblCumul.syc)

(Utilisation du fichier de commande CorrecRuptMult_ReseauSecond.R)

(Utilisation du fichier de commande Correct.syc)

Fin de la vague 2 à la fin de la liste des stations. On peut relancer une dernière fois CumRes2.syc pour un affichage général final et on peut également mettre à jour DblCumul.syc avec le dernier nom de variable homogénéisée.

Dans notre cas concret, 26 étapes supplémentaires ont été nécessaires pour corriger 108 ruptures d'homogénéité sur ce second réseau.

On pourrait s'étonner d'un nombre de ruptures bien plus important (par station) pour l'ETP que pour les températures, surtout que l'ETP est un paramètre météorologique beaucoup moins variant. Ceci s'explique cependant par le fait que l'ETP est un paramètre composite qui utilise le vent, l'humidité et le rayonnement, en plus de la température. Les ruptures d'homogénéité de chacune des variables constitutives tendent à s'ajouter.

On pourrait maintenant passer à l'homogénéisation de la variance. C'est ce qui a été tenté de plusieurs manières, cependant toutes ces tentatives ont abouti à des absurdités, comme des valeurs négatives ou un seuil bas différent de zéro. Donc cette étape s'est en fait résumée à une détection de l'homogénéité au sens de la variance. Par simplicité et reproductibilité, nous avons repris les procédures d'homogénéisation de la variance mais avec une homogénéisation nulle.

3.2.2.2. Homogénéisation de la variance

Une bonne partie du travail de préparation de l'homogénéisation est à faire. Il faut d'abord conserver une trace synthétique du travail exécuté sur la moyenne et préparer les données c'est-à-dire calculer les variances mensuelles des séries déjà homogénéisées d'un point de vue de la moyenne.

Pour le premier point, on peut confectionner directement le fichier de configuration des ruptures de variance (ConfigRuptV.ods si c'est le tableur Calc de LibreOffice) avec les 3 colonnes comme ceci

Station	NomVarHm	EtapeHV
Vichy-Charmeil	ETP03060H04	X
Lurcy-Levis	ETP03155H06	X
St-Nicolas	ETP03248H07	X
Colombier	ETP07068H08	X
Millau	ETP12145H03	X
Aubrac	ETP12214H09	X
Salles la Source	ETP12254H10	X
Villefranche de R.	ETP12300H11	X
Coltines	ETP15053H12	X
Marcenat	ETP15114H13	X
Maurs	ETP15122H14	X
Brive	ETP19031H15	X
Naves	ETP19146H16	X
Ussel les Plaines	ETP192754H17	X
Ahun	ETP23001H18	X
Bourganeuf	ETP23030H19	X
St Etienne Bout.	ETP42005H01	X
Perreux	ETP42170H20	X
Le Puy-Loudes	ETP43062H21	X
Fontannes	ETP43096H22	X
Landos-Charbon	ETP43111H23	X
Mazet-Volamont	ETP43130H24	X
Saugues	ETP43234H25	X
Gourdon	ETP46127H02	X
Le Montat	ETP46197H26	X
Chateau-Chinon	ETP58062H27	X
Ambert	ETP63003H28	X
St Genès	ETP63345H29	X
St Gervais	ETP63354	X
Montredon	ETP81182H30	X
Limoges-Bellegarde	ETP870856H05	X
St Yriex la Perche	ETP87187H31	X

Pour le second point, on utilise un fichier de commande SYSTAT/MYSTAT

(Utilisation du fichier de commande VarMens.sync, à confectionner entièrement ou à adapter d'une précédente utilisation)

```

REM*****
REM Programme de fabrication d'écart-type mensuel *
REM*****

REM Variable ETP

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\ETPMfH.syd"

REM Création variable d'indice
let AnMois=Int(Date/100)

REM Transformation des variables pour les rendre plus universelles

let ETP03060Hm=ETP03060H04 (écriture manuelle ou reprise depuis ConfigRuptV.ods)
let ETP03155Hm=ETP03155H06
let ETP03248Hm=ETP03248H07
let ETP07068Hm=ETP07068H08
let ETP12145Hm=ETP12145H03
let ETP12214Hm=ETP12214H09
let ETP12254Hm=ETP12254H10
let ETP12300Hm=ETP12300H11
let ETP15053Hm=ETP15053H12
let ETP15114Hm=ETP15114H13
let ETP15122Hm=ETP15122H14
let ETP19031Hm=ETP19031H15
let ETP19146Hm=ETP19146H16
let ETP192754Hm=ETP192754H17
let ETP23001Hm=ETP23001H18
let ETP23030Hm=ETP23030H19
let ETP42005Hm=ETP42005H01
let ETP42170Hm=ETP42170H20
let ETP43062Hm=ETP43062H21
let ETP43096Hm=ETP43096H22
let ETP43111Hm=ETP43111H23
let ETP43130Hm=ETP43130H24
let ETP43234Hm=ETP43234H25
let ETP46127Hm=ETP46127H02
let ETP46197Hm=ETP46197H26
let ETP58062Hm=ETP58062H27
let ETP63003Hm=ETP63003H28
let ETP63345Hm=ETP63345H29
let ETP63354Hm=ETP63354
let ETP81182Hm=ETP81182H30
let ETP870856Hm=ETP870856H05
let ETP87187Hm=ETP87187H31

REM Fabrication des écart-types et stockages

BY AnMois

SAVE 'C:\EspaceSYSTAT\CA23\Homogeneisation\ETPMfvm.syd'
(même nom de fichier que données quotidiennes mais avec suffixe vm pour variance mensuelle)
STATS
STATS ETP03060Hm ETP03155Hm ETP03248Hm ETP07068Hm ETP12145Hm ETP12214Hm ETP12254Hm,
ETP12300Hm ETP15053Hm ETP15114Hm ETP15122Hm ETP19031Hm ETP19146Hm ETP192754Hm,
ETP23001Hm ETP23030Hm ETP42005Hm ETP42170Hm ETP43062Hm ETP43096Hm ETP43111Hm,
ETP43130Hm ETP43234Hm ETP46127Hm ETP46197Hm ETP58062Hm ETP63003Hm ETP63345Hm,
ETP63354Hm ETP81182Hm ETP870856Hm ETP87187Hm / SD

RUN

```

De fait l'homogénéisation de la variance est déjà terminée, donc il ne reste plus qu'à effectuer quelques tâches du transfert de l'homogénéisation de la variance vers les données quotidiennes.

3.2.2.3. Transfert de l'homogénéisation de la variance vers les données quotidiennes

En préambule, il faut que le fichier « ConfigRuptV.syd » soit exporté au format .csv, en exigeant que chaque texte soit entouré de guillemets. Pour information, voici ce que les premières lignes de « ConfigRuptV.csv » doivent donner :

```

"Station","NomVarHm","EtapeHV"
"Vichy-Charmeil","ETP03060H04","X"
"Lurcy-Levis","ETP03155H06","X"
...

```

Il faut également exporter au format .txt le fichier SYSTAT/MYSTAT de la fin d'homogénéisation de la moyenne, soit « ETPMfH.syd » vers « ETPMfH.txt » dans notre exemple.

Ensuite, on transférera les 2 fichiers « ConfigRuptV.csv » et « ETPMfH.txt » sous l'environnement R. Attention, il est impératif de joindre aussi au moins un fichier TabRupt_Etape01.txt. Or, ici, il y a eu zéro étape d'homogénéisation de la variance. Donc, on en créera un minimaliste avec "Obs" en ligne 1 et le nombre de mois en ligne 2.

(Utilisation du fichier de commande AppliCorVarDonnéesQ.R)

Cette procédure fournit en sortie le fichier « ListRuptV.txt » qui contient l'ensemble des périodes homogènes au sens de la variance et les corrections appliquées, soit ici une correction multiplicative par un facteur 1 partout. La procédure fournit surtout le fichier des séries Q homogénéisées au sens de la variance (et de la moyenne) soit « ETPMfHV.txt » dans notre exemple, avec un nommage et une forme adaptée à l'usage des procédures de contrôle d'homogénéisation, que nous souhaitons précisément utiliser.

Ces 2 fichiers sont ensuite transférés sous l'environnement SYSTAT/MYSTAT où le fichier des séries homogénéisées est converti au format du logiciel de statistiques soit « ETPMfHV.syd » dans notre exemple.

On lance les vérifications de bon traitement.

On vérifie d'abord que les séries homogénéisées finales sont toujours homogènes au sens de la moyenne. Pour cela, on relance un processus d'homogénéisation de la moyenne.

(Utilisation du fichier de commande CumResTest.syc)

Ce fichier traite de l'ensemble du réseau (avec et sans manque au départ).

```
REM Paramètre ETP
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM Création de la référence spécifique à cette méthode par sous-réseau sans manque
USE "ETPMfHV.syd"
LET ETPRegHmv=AVG(ETP03060Hmv,ETP12145Hmv,ETP42005Hmv,ETP46127Hmv,ETP870856Hmv)
SAVE "ETPMfHV.syd"
RUN

USE "ETPMfHV.syd"
REM Station n°1 Vichy-Charmeil
LET X=ETPRegHmv
LET Y=ETP03060Hmv
RUN
...
```

Afficher l'ensemble des graphiques et vérifier qu'ils ont peu évolué par rapport à la fin du processus d'homogénéisation de la moyenne. S'il y a évolution sensible et qu'une nouvelle rupture est apparue, c'est très probablement dû à une erreur de manipulation.

Si le test est probant, on relance l'homogénéisation de la variance.

(Utilisation du fichier de commande VarMens.syc)

Attention n'exécuter que la dernière partie, relative aux vérifications.

```
REM Pour vérification de la non-altération de l'homogénéisation lors du processus de correction.
USE "C:\EspaceSYSTAT\CA23\Homogeneisation\ETPMfHV.syd"

REM Création variable d'indice
LET AnMois=Int(Date/100)

BY AnMois
```

```
SAVE 'C:\EspaceSYSTAT\CA23\Homogeneisation\ETPMfvmTest.syd'
STATS
STATS ETP03060Hmv ETP03155Hmv ETP03248Hmv ETP07068Hmv ETP12145Hmv ETP12214Hmv ETP12254Hmv,
ETP12300Hmv ETP15053Hmv ETP15114Hmv ETP15122Hmv ETP19031Hmv ETP19146Hmv ETP192754Hmv,
ETP23001Hmv ETP23030Hmv ETP42005Hmv ETP42170Hmv ETP43062Hmv ETP43096Hmv ETP43111Hmv,
ETP43130Hmv ETP43234Hmv ETP46127Hmv ETP46197Hmv ETP58062Hmv ETP63003Hmv ETP63345Hmv,
ETP63354Hmv ETP81182Hmv ETP870856Hmv ETP87187Hmv / SD
```

RUN

(Utilisation du fichier de commande CumResVTest.syc)

Ce fichier traite de l'ensemble du réseau (avec et sans manque au départ).

REM Paramètre ETP

FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM Création de la référence spécifique à cette méthode par sous-réseau sans manque

USE "ETPMfvmTest.syd"

LET ETPRegHmv=AVG(ETP03060Hmv,ETP12145Hmv,ETP42005Hmv,ETP46127Hmv,ETP870856Hmv)

SAVE "ETPMfvmTest.syd"

RUN

USE "ETPMfvmTest.syd"

REM Station n°1 Vichy-Charmeil

LET X=ETPRegHmv

LET Y=ETP03060Hmv

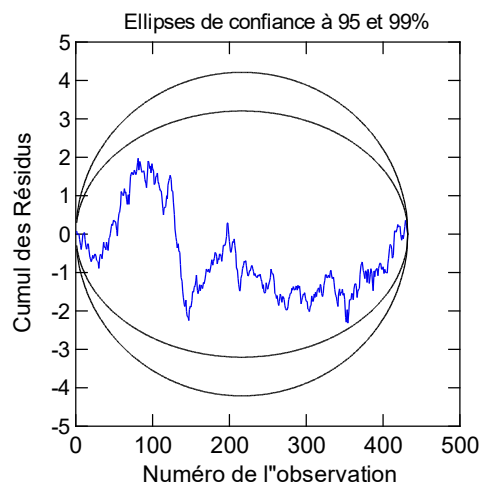
RUN

...

Bien afficher l'ensemble des graphiques mais ici le travail consiste à vérifier si les séries sont bien homogènes en variance. Pour cela, nous avons construit une règle empirique qui sépare les séries en 3 catégories.

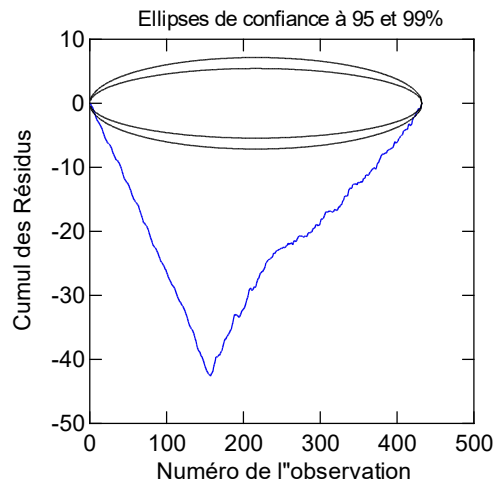
La catégorie 1 correspond aux séries homogènes en variance, dont les résidus de régression sont situés dans les ellipses de confiance, ou peu s'en faut si on tient compte que les véritables ellipses sont un peu plus larges. En voici l'exemple de St Etienne-Bouthéon :

Sigma Mens. ETP St Etienne-Bouthéon Hmv 1980-2015



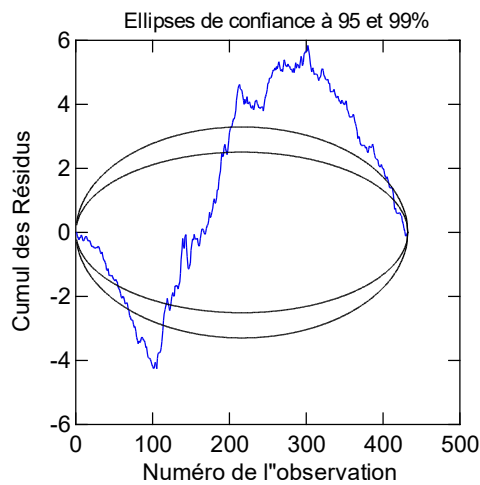
La catégorie 3 correspond aux séries qui sont fortement hétérogènes, comme par exemple celle de St-Nicolas :

Sigma Mens. ETP St Nicolas Hmv 1980-2015



Enfin, la catégorie 2 est définie indirectement par rapport aux deux autres, comme Villefranche de Rouergue, pour vérifier que le traitement retenu pour la catégorie 1, puisse y être étendu, sans trop de perturbations :

Sigma Mens. ETP Villefranche de R. Hmv 1980-2015



Au final, on classe toutes les séries dans les catégories suivantes :

Station	Série	Catégorie
Vichy-Charmeil	ETP03060Hmv	1
Lurcy-Levis	ETP03155Hmv	1
St-Nicolas	ETP03248Hmv	3
Colombier	ETP07068Hmv	3
Millau	ETP12145Hmv	2
Aubrac	ETP12214Hmv	3
Salles la Source	ETP12254Hmv	3
Villefranche de R.	ETP12300Hmv	2
Coltines	ETP15053Hmv	3
Marcenat	ETP15114Hmv	3
Maurs	ETP15122Hmv	3
Brive	ETP19031Hmv	1
Naves	ETP19146Hmv	3
Ussel les Plaines	ETP192754Hmv	1
Ahun	ETP23001Hmv	3
Bourganeuf	ETP23030Hmv	3
St Etienne Bout.	ETP42005Hmv	1
Perreux	ETP42170Hmv	2
Le Puy-Loudes	ETP43062Hmv	1
Fontannes	ETP43096Hmv	3
Landos-Charbon	ETP43111Hmv	3
Mazet-Volamont	ETP43130Hmv	3
Saugues	ETP43234Hmv	3
Gourdon	ETP46127Hmv	1
Le Montat	ETP46197Hmv	1
Chateau-Chinon	ETP58062Hmv	3
Ambert	ETP63003Hmv	3
St Genès	ETP63345Hmv	3
St Gervais	ETP63354Hmv	3
Montredon	ETP81182Hmv	3
Limoges-Bellegarde	ETP870856Hmv	1
St Yriex la Perche	ETP87187Hmv	1

Le choix pourrait être de ne poursuivre le travail qu'avec les séries dont les variances sont homogènes. Cependant, on perdrait alors environ les 2/3 du réseau qui est déjà particulièrement lâche concernant l'ETP. Après de nombreux essais, nous avons abouti à la conception d'une forme d'homogénéisation de la variabilité par l'intermédiaire de l'analyse de l'organisation chronologique décrite après l'homogénéisation. Ceci permet de conserver l'ensemble du réseau.

On aborde maintenant l'homogénéisation des précipitations.

3.2.3. Précipitations

Les exemples sont choisis dans ce qui a été fait pour la zone Nord du Massif central. On déroulera l'homogénéisation au sens de la moyenne puis au sens de la variance. Des détails seront donnés concernant les opérations spécifiques à cette variable. Pour les autres, prière de se reporter aux chapitres sur les températures ou sur l'ETP.

3.2.3.1. Homogénéisation de la moyenne

Étape 0 : Calcul initial de la référence (*Utilisation du fichier de commande Correct.sys, il se confectionne étape par étape*)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM On s'occupe des RR
REM*****

REM*****
REM Etape 0 *
REM*****

USE "RRMnc.syd"

REM Calcul initial de la référence d'homogénéisation
LET RRReg=AVG(RR03060,RR23001,RR23013,RR23021,RR23025,RR23046,RR23067,RR23069,RR23075,RR230961,RR23114,
RR23142,RR23144,RR23165,RR42005,RR42163,RR42168,RR42299,RR58062,RR69203,RR71472,RR89349)
(il y avait donc 22 stations sans manque dans ce réseau)
RUN
SAVE "RRMNH.syd"
RUN
```

3.2.3.1.1. Vague 1, réseau sans valeur manquante

On rappelle qu'on recalculera la référence à chaque étape, pour ce réseau.

Étape 1 : Recherche et correction de la station comportant la rupture la plus flagrante.

(Utilisation du fichier de commande CumRes1.sys, à confectionner ou adapter avant 1^{er} lancement)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM La méthode sépare les stations en deux groupes et produit donc l'homogénéisation en deux
REM grandes vagues. La première vague voit le traitement des 22 stations n'ayant aucune valeur
REM manquante sur la période 1980-2015

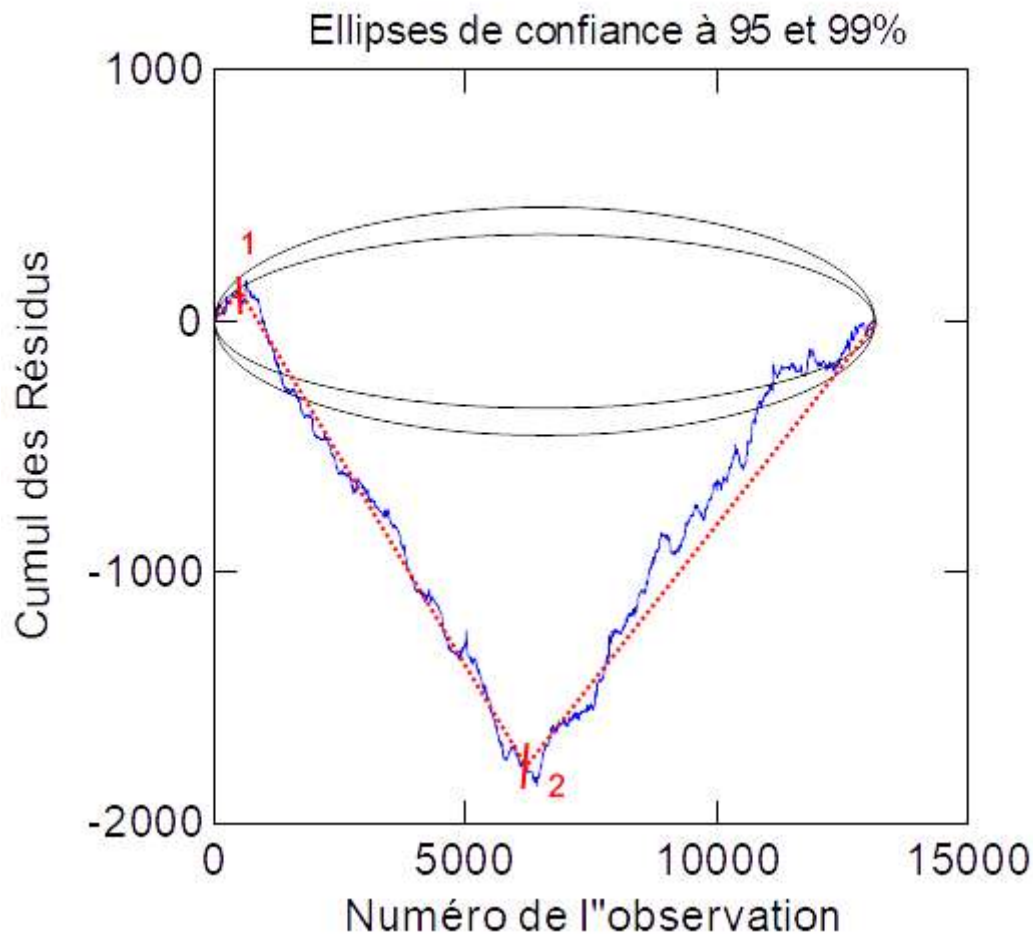
USE "RRMNH.syd"
REM Station n°1 Vichy-Charmeil
LET X=RRReg
LET Y=RR03060
RUN
...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
XMIN=0 YLABEL='Cumul des Résidus' TITLE='RR Vichy-Charmeil H00 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "RRQ03VichyC" /WMF
...
```

On consulte l'ensemble des graphiques qui ont été produits par ce fichier de commande et on repère celui dont les évolutions s'écartent le plus des ellipses de confiance d'homogénéité.

La station de La Courtine est la plus hétérogène. On se contraint à ré-exécuter la portion de CumulRes1.sys qui est relative à cette station et on affiche ou imprime le graphe en grand format :

RR La Courtine H00 1980-2015



On repère approximativement des ruptures en positions 500 et 6000.

(Utilisation du fichier de commande RegresRuptures.syc)

On produit le fichier de coefficients « RuptParam.txt ».

(Utilisation du fichier de commande DoubleCumul.syc)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"  
USE "RRMNH.syd"  
...  
LET X=RRReg  
LET Y=RR23067H  
LET NumObs=Case  
LET Decal0=0 remarquer l'absence d'usage du décalage du zéro car les précipitations ne sont jamais négatives  
...
```

Il faut exécuter uniquement la portion relative à la station considérée (La Courtine INSEE 23067). Cette procédure fournit les fichiers « DblCumul.txt » et « FichY.txt ».

Les fichiers « RuptParam.txt », « DblCumul.txt » et « FichY.txt » sont envoyés dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptMult_ReseauBase.R)

Veillez noter qu'on utilise pour les précipitations une correction multiplicative dès l'homogénéisation de la moyenne, ce qui est une évidence vu la nature cumulative de la variable.

```
# Script de correction de ruptures d'homogénéité, par coefficient multiplicatif, réseau de base
#####

# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation, nombre de stations du réseau de référence

Chemin<-"~/home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-1
NumEtape<-"Etape01"
NbStatRef<-22

...

```

Cette procédure produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt », qui sont à retransmettre dans l'environnement SYSTAT/MYSTAT. Le fichier « TabRupt_Etapexx.txt » est à contrôler et à stocker dans un répertoire spécifique. Le fichier « YCorr.txt » rejoint le répertoire de travail pour intégration au fichier de l'ensemble des données à homogénéiser (RRMNH.syd dans notre cas).

(Utilisation du fichier de commande Correct.syc)

```
REM*****
REM Etape 1 *
REM*****
REM Station de La Courtine

IMPORT "YCorr.txt" / TYPE=ASCII
LET RR23067H01=YCorr
ESAVE 'YCorr.syd'
MERGE 'RRMNH.syd' 'YCorr.syd' (RR23067H01)

REM On fait le recalcul du vecteur régional
LET RRReg01=AVG (RR03060,RR23001,RR23013,RR23021,RR23025,RR23046,RR23067H01,RR23069,RR23075,RR230961,RR23114,
                RR23142,RR23144,RR23165,RR42005,RR42163,RR42168,RR42299,RR58062,RR69203,RR71472,RR89349)

RUN
SAVE "RRMNH.syd"
RUN

```

Étapes 2 à nn : Recherche et correction de la station comportant la rupture la plus flagrante.

(Utilisation du fichier de commande CumRes1.syc)

(Utilisation du fichier de commande RegresRuptures.syc)

(Utilisation du fichier de commande DblCumul.syc)

(Utilisation du fichier de commande CorrecRuptMult_ReseauBase.R)

(Utilisation du fichier de commande Correct.syc)

Fin de la vague 1 quand les graphes de CumRes1.syc sont tous dans les ellipses de confiance. Il faut pour cela le vérifier en exécutant CumRes1.syc de manière isolée, une dernière fois.

Dans notre cas concret, 8 étapes ont été nécessaires pour corriger 12 ruptures d'homogénéité sur ce réseau.

3.2.3.1.2. Vague 2, autres stations

On rappelle que la référence n'évoluera plus, pour ce réseau. Le processus ressemble beaucoup au précédent mais avec des noms de fichiers de commande parfois distincts.

Étape nn+1: Recherche et correction de ruptures d'homogénéité, station par station. L'ordre importe peu.

(Utilisation du fichier de commande CumRes2.syc, à confectionner ou adapter avant 1^{er} lancement)

```
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"

REM Cette méthode sépare les stations en deux groupes et produit donc l'homogénéisation en deux
REM grandes vagues. La deuxième vague voit le traitement des 38 stations ayant des valeurs
REM manquantes. Le "vecteur régional" est celui obtenu précédemment et il reste fixe.

REM Deuxième vague d'homogénéisation. Vecteur régional fixe RRReg08.

USE "RRMNH.syd"
REM Station n°2 Lurcy-Levis
LET X=RRReg08
LET Y=RR03155
RUN
...

REM tracé avec ajout d'un sous-titre
BEGIN
PLOT CumRes Ellip99P Ellip95P Ellip95N Ellip99N*NumObs /OVERLAY LINE XLABEL='Numéro de l'observation',
      XMIN=0 YLABEL='Cumul des Résidus' TITLE='RR Lurcy-Levis H08 1980-2015' COLOR=2,10,10,10,10 LEGEND=NONE ,
      SIZE = 0,0,0,0,0
WRITE '\>Ellipses de confiance à 95 et 99%' /LOC=0.5,4
END
GSAVE "RRQ03LurcyL" /WMF
...
```

On affiche uniquement le graphe de la station considérée. Ici, il s'agit de Lurcy-Levis (INSEE 03155), pour laquelle on repère des ruptures probables aux positions approximatives 4000, et 9000.

On produit le fichier « RuptParam.txt ».

(Utilisation du fichier de commande DblCumul.syc)

On n'oubliera pas de réactualiser une dernière fois le nom de la variable de référence, puisque ça n'a pas été fait à l'issue de la dernière étape de la vague 1.

On produit les doubles cumuls et la variable banalisée Y à corriger (le cumul de précipitations de Lurcy-Levis en ce cas). On joint les fichiers « DblCumul.txt » et « FichY.txt » au fichier « RuptParam.txt » obtenu juste avant et on les transfère dans l'environnement R.

(Utilisation du fichier de commande CorrecRuptMult_ReseauSecond.R)

```
# Script de correction de ruptures d'homogénéité, par coefficient multiplicatif, réseau secondaire
#####

# Initialisations, chemin, la référence est t'elle la dernière période, sinon inscription manuelle
# Numéro de l'étape d'homogénéisation

Chemin<-"/home/debian/Documents/Rfiles/Homogénéisation/"
RefPerFinale<-TRUE
PerRef<-2
NumEtape<-"Etape08" (bien mettre les numéros d'étape dans la continuité du réseau sans manque)

« IDEM »
```

L'utilisation en est strictement équivalente à celle du réseau sans manque, sauf qu'on n'a plus à renseigner la taille du réseau.

On produit les fichiers « TabRupt_Etapexx.txt » et « YCorr.txt »

(Utilisation du fichier de commande Correct.syc)

```
REM*****  
REM Etape 9 *  
REM*****  
REM Station de Lurcy-Levis  
  
REM Attention, on ne change plus la référence donc pas de sur-correction ni de recalcul.  
  
IMPORT "YCorr.txt" / TYPE=ASCII  
LET RR03155H09=YCorr  
ESAVE 'YCorr.syd'  
MERGE 'RRMNH.syd' 'YCorr.syd' (RR03155H09)  
  
RUN  
SAVE "RRMNH.syd"  
RUN
```

Étapes nn+2 à pp : Recherche et correction des ruptures d'homogénéité de la station suivante.

(Utilisation du fichier de commande CumRes2.syc)

On affichera également le cumul des résidus sur la station de l'étape précédente pour vérification de la bonne correction.

(Utilisation du fichier de commande RegresRuptures.syc)

(Utilisation du fichier de commande DblCumul.syc)

(Utilisation du fichier de commande CorrecRuptMult_ReseauSecond.R)

(Utilisation du fichier de commande Correct.syc)

Fin de la vague 2 à la fin de la liste des stations. On peut relancer une dernière fois CumRes2.syc pour un affichage général final et on peut également mettre à jour DblCumul.syc avec le dernier nom de variable homogénéisée.

Dans notre cas concret, 34 étapes supplémentaires ont été nécessaires pour corriger 67 ruptures d'homogénéité sur ce second réseau.

Les tentatives d'homogénéisation effective de la variance ont abouti aux mêmes types d'absurdités que pour l'ETP. Donc ce processus se résume également à une détection de l'homogénéité au sens de la variance.

3.2.3.2. Homogénéisation de la variance

On conserve d'abord une trace du travail exécuté sur la moyenne en confectionnant directement le fichier de configuration des ruptures de variance « ConfigRuptV.ods » avec 3 colonnes, dont voici un extrait :

Station	NomVarHm	EtapeHV
Vichy-Charneil	RR03060	X
Lurcy-Levis	RR03155H09	X
Marcillat	RR03161H10	X
...
St Prix	RR71472	X
St Symphorien	RR71482	X
St Léger Vauban	RR89349H08	X

On calcule les variances mensuelles des séries homogénéisées au sens de la moyenne.

(Utilisation du fichier de commande VarMens.sys, à confectionner entièrement ou à adapter d'une précédente utilisation)

```
REM*****
REM Programme de fabrication d'écart-type mensuel      *
REM*****

REM Variable RR

USE "C:\EspaceSYSTAT\CA23\Homogeneisation\RRMNH.sysd"

REM Création variable d'indice
let AnMois=Int(Date/100)

REM Transformation des variables pour les rendre plus universelles

let RR03060Hm=RR03060 (écriture manuelle ou reprise depuis ConfigRuptV.ods)
let RR03155Hm=RR03155H09
let RR03161Hm=RR03161H10
...

let RR71472Hm=RR71472
let RR71482Hm=RR71482
let RR89349Hm=RR89349H08

REM Fabrication des écart-types et stockages

BY AnMois

SAVE 'C:\EspaceSYSTAT\CA23\Homogeneisation\RRMnm.sysd'
STATS
STATS RR03060Hm RR03155Hm RR03161Hm RR03180Hm RR03248Hm RR03285Hm RR21445Hm,
RR21584Hm RR23001Hm RR23004Hm RR23008Hm RR23013Hm RR23021Hm RR23025Hm,
RR23030Hm RR23031Hm RR23036Hm RR23045Hm RR23046Hm RR23061Hm RR23063Hm,
RR23067Hm RR23069Hm RR23075Hm RR23079Hm RR23081Hm RR23087Hm RR23090Hm,
RR23093Hm RR230961Hm RR230967Hm RR230968Hm RR23106Hm RR23114Hm RR23118Hm,
RR23142Hm RR23144Hm RR23155Hm RR23165Hm RR23174Hm RR23176Hm RR23223Hm,
RR42005Hm RR42019Hm RR42163Hm RR42168Hm RR42170Hm RR42299Hm RR58062Hm,
RR58145Hm RR58199Hm RR63003Hm RR63038Hm RR63345Hm RR63354Hm RR69107Hm,
RR69203Hm RR71472Hm RR71482Hm RR89349Hm / SD

RUN
```

De fait l'homogénéisation de la variance est déjà terminée, donc il ne reste plus qu'à effectuer quelques tâches du transfert de l'homogénéisation de la variance vers les données quotidiennes.

3.2.3.3. Transfert de l'homogénéisation de la variance vers les données quotidiennes

En préambule, il faut que le fichier « ConfigRuptV.sysd » soit exporté au format .csv, en exigeant que chaque texte soit entouré de guillemets. Pour information, voici ce que les premières lignes de « ConfigRuptV.csv » doivent donner :

```
"Station","NomVarHm","EtapeHV"
"Vichy-Charmeil","RR03060","X"
"Lurcy-Levis","RR03155H09","X"
...
```

Il faut également exporter au format .txt le fichier SYSTAT/MYSTAT de la fin d'homogénéisation de la moyenne, soit « RRMNH.sysd » vers « RRMNH.txt » dans notre exemple.

Ensuite, on transférera les 2 fichiers « ConfigRuptV.csv » et « RRMNH.txt » sous l'environnement R. Attention, il est impératif de joindre aussi au moins un fichier TabRupt_Etape01.txt obtenu lors d'un processus d'homogénéisation de la variance. Or, ici, il n'y a eu aucune étape d'homogénéisation de la variance. On pourra prendre à la place un fichier « TabRupt_Etape01.txt » minimaliste déjà décrit (« Obs » en ligne 1 et le nombre de mois en ligne2).

(Utilisation du fichier de commande AppliCorVarDonnéesQ.R)

Cette procédure fournit en sortie le fichier « ListRuptV.txt » qui contient l'ensemble des périodes homogènes au sens de la variance et les corrections appliquées, soit ici une correction multiplicative qui vaut 1 partout. La procédure fournit surtout le fichier des séries Q homogénéisées au sens de la variance (et de la moyenne) soit « RRMNHV.txt » dans notre exemple, avec un nommage et une forme adaptés à l'usage des procédures de contrôle d'homogénéisation, que nous utiliserons dans la foulée.

Ces 2 fichiers sont ensuite transférés sous l'environnement SYSTAT/MYSTAT où le fichier des séries homogénéisées est converti au format du logiciel de statistiques soit « RRMNHV.syd » dans notre exemple.

On lance les vérifications de bon traitement.

On doit d'abord vérifier que les séries homogénéisées finales sont toujours homogènes au sens de la moyenne. On relance un processus d'homogénéisation de la moyenne.

(Utilisation du fichier de commande CumResTest.syc)

Ce fichier traite de l'ensemble du réseau (avec et sans manque au départ).

```
REM Paramètre RR
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"
REM Création de la référence spécifique à cette méthode par sous-réseau sans manque
USE "RRMNHV.syd"
LET RRRegHmv=AVG(RR03060Hmv, RR23001Hmv, RR23013Hmv, RR23021Hmv, RR23025Hmv, RR23046Hmv, RR23067Hmv,
RR23069Hmv, RR23075Hmv, RR230961Hmv, RR23114Hmv, RR23142Hmv, RR23144Hmv, RR23165Hmv,
RR42005Hmv, RR42163Hmv, RR42168Hmv, RR42299Hmv, RR58062Hmv, RR69203Hmv, RR71472Hmv, RR89349Hmv)
SAVE "RRMNHV.syd"
RUN

USE "RRMNHV.syd"
REM Station n°1 Vichy-Charmeil
LET X=RRRegHmv
LET Y=RR03060Hmv
RUN
...
```

On affiche l'ensemble des graphiques et on vérifie qu'ils n'ont pas évolué par rapport à la fin du processus d'homogénéisation de la moyenne. S'il y a évolution sensible, c'est très probablement dû à une erreur de manipulation.

Si le test est probant, on relance l'homogénéisation de la variance.

(Utilisation du fichier de commande VarMens.syc)

Attention on n'exécute que la dernière partie, relative aux vérifications.

```
REM Pour vérification de la non-altération de l'homogénéisation lors du processus de correction.
USE "C:\EspaceSYSTAT\CA23\Homogeneisation\RRMNHV.syd"

REM Création variable d'indice
LET AnMois=Int(Date/100)

BY AnMois

SAVE 'C:\EspaceSYSTAT\CA23\Homogeneisation\RRMnmTest.syd'
STATS
STATS RR03060Hmv RR03155Hmv RR03161Hmv RR03180Hmv RR03248Hmv RR03285Hmv RR21445Hmv,
RR21584Hmv RR23001Hmv RR23004Hmv RR23008Hmv RR23013Hmv RR23021Hmv RR23025Hmv,
RR23030Hmv RR23031Hmv RR23036Hmv RR23045Hmv RR23046Hmv RR23061Hmv RR23063Hmv,
RR23067Hmv RR23069Hmv RR23075Hmv RR23079Hmv RR23081Hmv RR23087Hmv RR23090Hmv,
RR23093Hmv RR230961Hmv RR230967Hmv RR230968Hmv RR23106Hmv RR23114Hmv RR23118Hmv,
RR23142Hmv RR23144Hmv RR23155Hmv RR23165Hmv RR23174Hmv RR23176Hmv RR23223Hmv,
RR42005Hmv RR42019Hmv RR42163Hmv RR42168Hmv RR42170Hmv RR42299Hmv RR58062Hmv,
RR58145Hmv RR58199Hmv RR63003Hmv RR63038Hmv RR63345Hmv RR63354Hmv RR69107Hmv,
RR69203Hmv RR71472Hmv RR71482Hmv RR89349Hmv / SD

RUN
```

(Utilisation du fichier de commande CumResVTest.syc)

Ce fichier traite de l'ensemble du réseau (avec et sans manque au départ).

```
REM Paramètre RR
FPATH "C:\EspaceSYSTAT\CA23\Homogeneisation"
REM Création de la référence spécifique à cette méthode par sous-réseau sans manque
USE "RRMnmTest.syd"
LET RRRegHmv=AVG(RR03060Hmv, RR23001Hmv, RR23013Hmv, RR23021Hmv, RR23025Hmv, RR23046Hmv, RR23067Hmv,
                RR23069Hmv, RR23075Hmv, RR230961Hmv, RR23114Hmv, RR23142Hmv, RR23144Hmv, RR23165Hmv,
                RR42005Hmv, RR42163Hmv, RR42168Hmv, RR42299Hmv, RR58062Hmv, RR69203Hmv, RR71472Hmv, RR89349Hmv)
SAVE "RRMnmTest.syd"
RUN

USE "RRMnmTest.syd"
REM Station n°1 Vichy-Charmeil
LET X=RRRegHmv
LET Y=RR03060Hmv
RUN
...
```

On affiche l'ensemble des graphiques et on vérifie si les séries sont bien homogènes en variance. On utilise la même catégorisation que celle qui a été initiée avec les ETP, à savoir catégorie 1 pour série homogène en variance, catégorie 3 pour série très in-homogène et catégorie 2 intermédiaire.

Le résultat de ce classement, pour l'ensemble des 60 stations du réseau, est 50 stations en catégorie 1, 10 stations en catégorie 2 et aucune en catégorie 3. Ceci constitue une distribution bien plus favorable que celle relative aux ETP. On a néanmoins effectué la même opération d'homogénéisation de la variabilité par l'intermédiaire de l'analyse de l'organisation chronologique, dans la mesure où elle avait été déjà développée.

C'est ce processus d'analyse organisationnelle qui est décrit dans le chapitre suivant.

4. Analyse organisationnelle des séries de données

Sans étude ou compétence spécifiques, 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.

Des développements académiques assez détaillés figurent dans le rapport technique du projet « AP3C-Creuse » pages 28 à 137. Ici on n'en rappellera que quelques éléments avant de décrire plus spécifiquement l'usage d'une méthode automatisée simplifiée d'extraction des organisations chronologiques.

4.1. Déroulement exploratoire

A l'origine, cette démarche exploratoire a été effectuée avec des procédures SYSTAT/MYSTAT, mais elle est disponible désormais sous l'environnement R (via les versions semi-automatiques de la section 4.3.). L'intégralité des procédures figure dans l'annexe « Fichiers Analyse Organisation.ods ».

Le principe général est de séparer dans une série de données ce qui relève du caractère aléatoire de ce qui constitue un caractère organisé. L'expérience acquise lors du projet « AP3C-Creuse » et les essais effectués dans AP3C nous ont amenés à effectuer l'extraction progressive de différents types d'organisation.

On peut aussi considérer que cette extraction constitue un calcul de coordonnées sur une base de fonctions chronologiques (qui dépendent du temps).

Tout d'abord on s'occupera de l'organisation de long terme dans laquelle on range la tendance (pas nécessairement linéaire) ainsi que les cycles et les évolutions de cycles. Ensuite, on traitera l'organisation de court terme dans le cadre d'un schéma autorégressif.

Ces 2 types d'organisation se réfèrent à l'espérance de la série soit un niveau 1. Les niveaux supérieurs peuvent être considérés de manière extensive (organisation des quantiles de la distribution) ou de manière intensive (organisation de la variance en ordre 2, de la dissymétrie en ordre 3, de l'aplatissement en ordre 4...). C'est cette deuxième approche qui a été suivie, mais uniquement vis-à-vis de la variance puisqu'on avait prouvé avec « AP3C-Creuse » que les ordres supérieurs étaient peu organisés.

De manière exploratoire, on étudie manuellement les séries une par une et on identifie, parmi une base maximaliste, les fonctions chronologiques qui expliquent une proportion significative de variance (soit dont les coordonnées sont statistiquement significatives dans une régression multiple non-linéaire). Quand on a effectué ces études manuelles, on les regroupe pour repérer les fonctions qui ressortent « suffisamment souvent » en significativité. La règle empirique qui a été adoptée est la présence significative dans au moins 1/4 des analyses pour pouvoir faire partie de la base de fonctions opérationnelle sur laquelle seront calculées, de manière automatisée, les coordonnées.

On dispose ainsi d'une assez grande diversité de descriptions possibles mais en mettant de côté des fonctionnements très atypiques.

On décrit succinctement la manière dont on a obtenu les bases de fonctions opérationnelles pour chaque type d'organisation et chaque paramètre.

4.2. Détermination des bases de fonctions opérationnelles

4.2.1. Organisation de long terme

4.2.1.1. Températures

Les températures minimales T_n et maximales T_x quotidiennes sont traitées séparément. Suite à l'expérience de « AP3C-Creuse », la base de fonctions maximaliste pouvait être constituée par :

- Une tendance linéaire : $A*t+B$
- Un cycle sinusoïdal de 12 mois (365,25 jours) avec
 - Amplitude variable en tendance $(Amp1C+Amp1Tend*t)$
 - Déphasage variable en tendance $(Phas1C+Phas1Tend*t)$
- Un cycle sinusoïdal de 6 mois (182,625 jours) avec
 - Amplitude et déphasage fixes $Amp2, Phas2$

A noter que le temps t est le numéro du jour à partir de la première observation, soit le 01/01/1980 dans notre cas. La sélection des fonctions fréquemment significatives aboutit à conserver la totalité de la base maximaliste, soit 8 coordonnées par variable climatique.

4.2.1.2. ETP

Pour l'analyse manuelle, on ne considère que les stations qui sont de catégorie 1, c'est-à-dire homogène, en variance. La base maximaliste des fonctions chronologiques a été calquée sur celle des températures, soit :

- Une tendance linéaire : $A*t+B$
- Un cycle sinusoïdal de 12 mois (365,25 jours) avec
 - Amplitude variable en tendance $(Amp1C+Amp1Tend*t)$
 - Déphasage variable en tendance $(Phas1C+Phas1Tend*t)$
- Un cycle sinusoïdal de 6 mois (182,625 jours) avec
 - Amplitude et déphasage fixes $Amp2, Phas2$

La sélection des fonctions fréquemment significatives aboutit également à conserver la totalité de la base maximaliste, soit 8 coordonnées.

4.2.1.3. Précipitations

Après de nombreux essais, on a adopté, pour l'analyse de données (donc aussi pour les projections climatiques qui suivront), la même transformation de variable préalable que ce qui avait été conduit pour la reconstitution des données, à savoir $LRR=\log_{10}(RR+0,01+alea_uniforme[0;0,2])$.

C'est la même base de fonctions maximaliste que pour T_n , T_x et ETP qui a été testée, sur les stations de catégorie 1, donc homogènes, en variance. Cependant, ceci a abouti à très peu de coordonnées significatives et avec une grande incertitude. Il a donc été décidé de ne rien retenir directement en termes d'organisation de long terme mais d'utiliser la coévolution avec les organisations de long terme des températures, par l'intermédiaire de l'organisation de court terme. Autrement dit, le schéma d'organisation de court terme des précipitations ne se calculera pas, comme pour les températures et l'ETP, sur des séries dont l'organisation de long terme aurait été enlevée mais sur les séries complètes. Des explications plus détaillées seront données dans la section appropriée.

On peut donc passer à l'étude de l'organisation de court terme.

4.2.2. Organisation de court terme

Lors de l'étude « AP3C-Creuse », on était arrivé, pour la température, à un schéma autorégressif d'ordre 7 au maximum, dont les coefficients étaient variables en tendance et en organisations sinusoïdales fixes de 6 et 12 mois (le schéma « Variable Optimisé »). C'est cette base que nous allons reprendre pour la température et l'ETP.

4.2.2.1. Températures

Contrairement à l'analyse de long terme, les températures minimales T_n et maximales quotidiennes T_x sont ici traitées conjointement. Le but en est double.

Tout d'abord, on utilise l'information supplémentaire disponible dans la variable alterne. En effet, la température maximale du jour dépend effectivement de la température minimale du matin qui elle-même dépend à l'évidence de la température maximale de la veille. Mais également on assure une intrication lors de l'utilisation des analyses dans le processus de projection. Si on ne le faisait pas, les séries de T_n et de T_x évolueraient de manière indépendante, sans rappel, ce qui rendrait impossible l'usage de variables combinées, à commencer par la température moyenne $T_m=(T_n+T_x)/2$.

Si on reprend l'écriture $T_nSSorgO$ de « AP3C-Creuse » pour le résidu de l'extraction de l'organisation de long terme de T_n , le schéma d'organisation maximaliste de court terme peut s'écrire :

$$T_nSSorgO(t) = A_1 * T_nSSorgO(t-1) + A_2 \dots + A_7 * T_nSSorgO(t-7) \\ + B_1 * T_xSSorgO(t-1) + B_2 \dots + B_7 * T_xSSorgO(t-7)$$

avec $A_i = A_iC + A_iTend * t + A_iAmp1 * \sin(2\pi t / 365, 25 + A_iPhas1) + A_iAmp2 * \sin(2\pi t / 182, 625 + A_iPhas2)$
et B_i de même nature

Pour T_x , il peut s'écrire :

$$T_xSSorgO(t) = C_0 * T_nSSorgO(t) + C_1 * T_nSSorgO(t-1) + C_2 \dots + C_7 * T_nSSorgO(t-7) \\ + D_1 * T_xSSorgO(t-1) + D_2 \dots + D_7 * T_xSSorgO(t-7)$$

avec $C_i = C_iC + C_iTend * t + C_iAmp1 * \sin(2\pi t / 365, 25 + C_iPhas1) + C_iAmp2 * \sin(2\pi t / 182, 625 + C_iPhas2)$
et D_i de même nature

Une fois appliqué manuellement sur l'ensemble du réseau, on ne conserve que les fonctions « suffisamment significatives », comme pour l'organisation de long terme.

Il reste pour T_n : A_1 et A_2 complets, A_3C , A_5C et A_6C , B_1 et B_2 complets, B_3 sans la tendance, B_6C et B_7C , soit 34 coordonnées.

Il reste pour T_x : C_0 et C_1 complets, C_2 cycle de 6 mois, C_3 cycle de 6 mois, D_1 complet, D_2 sans tendance, D_3 cycle 6 mois, D_4C et D_6C , soit 32 coordonnées.

4.2.2.2. ETP

Si on veut pouvoir utiliser l'ETP conjointement avec les températures, il faut adopter le même principe d'intrication qu'entre les températures minimales et maximales seules. C'est-à-dire qu'il faut étudier des schémas d'organisation de court terme qui sont conjoints entre les 3 paramètres. On obtient deux bénéfices. Premièrement, les séries d'ETP qui seront produites à partir de ce schéma seront en concordance avec les séries de T_n et T_x . Deuxièmement, ce schéma offre la possibilité d'une seconde itération des projections de températures, sur les postes qui mesurent

l'ETP. Nous aurons donc les projections de température qui ont été obtenues seules et celles qui ont été obtenues conjointement avec l'ETP. Ceci donnera un embryon de possibilité de test de la robustesse des projections de nature statistique que nous produisons.

Les schémas d'organisation maximalistes de court terme peuvent donc s'écrire :

$$\begin{aligned} \text{ETPSSorgO}(t) = & A0 * \text{TnSSorgO}(t) + A1 * \text{TnSSorgO}(t-1) + A2 \dots + A7 * \text{TnSSorgO}(t-7) \\ & + B0 * \text{TxSSorgO}(t) + B1 * \text{TxSSorgO}(t-1) + B2 \dots + B7 * \text{TxSSorgO}(t-7) \\ & + C1 * \text{ETPSSorgO}(t-1) + C2 \dots + C7 * \text{ETPSSorgO}(t-7) \end{aligned}$$

avec $A_i = A_{iC} + A_{iTend} * t + A_{iAmp1} * \sin(2\pi t / 365, 25 + A_{iPhas1}) + A_{iAmp2} * \sin(2\pi t / 182, 625 + A_{iPhas2})$
 B_i et C_i de même nature

$$\begin{aligned} \text{TnSSorgO}(t) = & D1 * \text{TnSSorgO}(t-1) + D2 \dots + D7 * \text{TnSSorgO}(t-7) \\ & + E1 * \text{TxSSorgO}(t-1) + E2 \dots + E7 * \text{TxSSorgO}(t-7) \\ & + F1 * \text{ETPSSorgO}(t-1) + F2 \dots + F7 * \text{ETPSSorgO}(t-7) \end{aligned}$$

avec $D_i = D_{iC} + D_{iTend} * t + D_{iAmp1} * \sin(2\pi t / 365, 25 + D_{iPhas1}) + D_{iAmp2} * \sin(2\pi t / 182, 625 + D_{iPhas2})$
 E_i et F_i de même nature

$$\begin{aligned} \text{TxSSorgO}(t) = & G0 * \text{TnSSorgO}(t) + G1 * \text{TnSSorgO}(t-1) + G2 \dots + G7 * \text{TnSSorgO}(t-7) \\ & + H1 * \text{TxSSorgO}(t-1) + H2 \dots + H7 * \text{TxSSorgO}(t-7) \\ & + I1 * \text{ETPSSorgO}(t-1) + I2 \dots + I7 * \text{ETPSSorgO}(t-7) \end{aligned}$$

avec $G_i = G_{iC} + G_{iTend} * t + G_{iAmp1} * \sin(2\pi t / 365, 25 + G_{iPhas1}) + G_{iAmp2} * \sin(2\pi t / 182, 625 + G_{iPhas2})$
 H_i et I_i de même nature

Une fois appliqué manuellement sur l'ensemble du réseau, on ne conserve que les fonctions « suffisamment significatives ».

Il reste pour ETP : A0 et A1 complets, A5 et A6 cycles de 6 mois, B0 et B1 complets, B2 sans le cycle de 12 mois, B3 et B7 constantes et cycles de 12 mois, C1 et C2 sans tendances, C3C, C4C, C5C, C6C, C7 sans cycles, soit 54 coordonnées.

Il reste pour Tn : D1 complet, D2 constante et cycle de 12 mois, D3 constante et cycle de 6 mois, D6C, E1 complet, E2 constante et cycle 12 mois, E3 constante et cycle de 12 mois, B5 cycle de 12 mois, B7C, F1 et F2 complets, F5 cycle de 12 mois et F7 cycle de 6 mois, soit 44 coordonnées.

Il reste pour Tx : G0 et G1 complets, G2 cycles de 6 et 12 mois, H1 complet, H2 sans tendance, H3 constante et cycle 6 mois, I1 complet, I3C soit 37 coordonnées.

4.2.2.3. Précipitations

On rappelle la transformation de variable adoptée $LRR = \log_{10}(RR + 0,01 + \text{alea_uniforme}[0;0,2])$. De plus, le schéma d'organisation maximaliste s'articule sur les séries complètes de LRR, mais aussi de Tn et de Tx. Il s'écrit donc :

$$\begin{aligned} \text{LRR}(t) = & Cste + A0 * \text{Tn}(t) + A1 * \text{Tn}(t-1) + A2 \dots + A7 * \text{Tn}(t-7) \\ & + B0 * \text{Tx}(t) + B1 * \text{Tx}(t-1) + B2 \dots + B7 * \text{Tx}(t-7) \\ & + C1 * \text{LRR}(t-1) + C2 \dots + C7 * \text{LRR}(t-7) \end{aligned}$$

avec $A_i = A_{iC} + A_{iTend} * t + A_{iAmp1} * \sin(2\pi t / 365, 25 + A_{iPhas1}) + A_{iAmp2} * \sin(2\pi t / 182, 625 + A_{iPhas2})$
 B_i et C_i de même nature

Notez bien que pour une production de projections, ce genre de schéma nécessite un calcul préalable complet des températures (seules ou en combinaison avec l'ETP).

Il reste pour LRR : Cste, A0 complet, A1C, A3 cycle de 6 mois, B0 et B1 sans tendances, B2 constante et cycle de 12 mois, B4 cycle de 6 mois, C1 complet, C2 constante et cycle de 12 mois, C3C, C6 constante et cycle de 12 mois, soit 38 coordonnées.

Une fois l'organisation de court terme traitée, on s'occupe de l'organisation d'ordre 2 par l'intermédiaire de l'écart-type mensuel.

4.2.3. Organisation de la variance

Pour des raisons pratiques (même unité en particulier) l'étude de cette organisation se fait sur l'écart-type mensuel des séries dont on a ôté l'ensemble des organisations de long et de court termes. On a repris comme schéma maximaliste ce qui avait déjà été utilisé dans « AP3C-Creuse », à savoir :

- Une tendance linéaire : $A*t+B$
- Un cycle sinusoïdal de 12 mois (365,25 jours) avec
 - Amplitude variable en tendance $(Amp1C+Amp1Tend*t)$
 - Déphasage variable en tendance $(Phas1C+Phas1Tend*t)$
- Un cycle sinusoïdal de 6 mois (182,625 jours) avec
 - Amplitude et déphasage fixes $Amp2, Phas2$

Soit un maximum de 8 coordonnées par variable. Attention, ici le temps t est exprimé en mois (à partir du 01/1980 dans notre exemple) et il faudra une transformation de variable temporelle pour appliquer l'organisation de variance à des données quotidiennes.

4.2.3.1. Températures

Que ce soit pour les températures minimales ou les températures maximales, la base opérationnelle de fonctions chronologiques comprend la tendance linéaire, ainsi que les cycles de 6 et 12 mois fixes, soit 6 coordonnées par variable.

4.2.3.2. ETP

Bien se rappeler que les températures sont ré-examinées de concert avec l'ETP , car la désorganisation de l'étape précédente a été également menée conjointement. Pour celles-ci on retrouve la même base opérationnelle, avec 6 fonctions. Ceci n'était pas une obligation car certaines informations auraient pu transiter préférentiellement par l'ETP. D'ailleurs, les valeurs des coordonnées obtenues sur les fonctions chronologiques diffèrent parfois de manière significative de l'examen des températures seules.

Pour l'ETP strictement, le schéma opérationnel est plus complet puisqu'il inclut une tendance significative à la hausse de l'amplitude du cycle de 12 mois (de l'écart-type mensuel). Cela fait donc 7 coordonnées. De plus, ce schéma concerne la variance qui n'est homogène que sur une partie du réseau. On a donc séparé les 7 groupes de coordonnées selon l'appartenance à la catégorie 1 (homogène) ou non. On a ensuite vérifié si les sous-groupes se comportaient de manière différente, ce qui nécessiterait un bornage à la gamme de valeurs obtenues sur le sous-groupe des stations homogènes. Il se trouve que ce cas se produit uniquement pour la tendance de l'écart-type mensuel et un bornage à un maximum de +0,28°C par siècle a été retenu (pour l'application à l'extraction automatisée de l'organisation).

On peut rappeler que ce genre de bornage constitue une forme d'homogénéisation de la variance des séries, qui s'applique de fait plus aux projections (c'est-à-dire aux ré-analyses sur la période d'observation) qu'aux observations elles-mêmes.

4.2.3.3. Précipitations

La base opérationnelle de fonctions chronologiques est la même que celle des températures soit la tendance linéaire, ainsi que les cycles de 6 et 12 mois fixes, c'est-à-dire 6 coordonnées par variable. Il n'y a pas nécessité de bornage.

On a désormais tous les éléments pour passer à l'extraction automatisée des organisations.

4.3. Extraction automatisée des organisations

Toutes les procédures sont disponibles dans l'environnement R. Pour chacune, on dispose d'une version entièrement automatisée et d'une version semi-automatique dont la confection a été rendue indispensable pour pallier à l'instabilité de certains processus de régression multiple (jusqu'à 54 coefficients-coordonnées à calculer!). Cette version permet également de refaire une démarche exploratoire sur d'autres zones que le Massif central ou sur d'autres périodes, pour lesquelles la base opérationnelle pourrait être différente.

On présentera d'abord l'usage de la version semi-automatique puis celle de la version automatique.

4.3.1. Températures

Comme précédemment, on va prendre comme exemple le cas de la zone Nord du Massif central. Il faut tout d'abord importer dans l'environnement R les deux fichiers « TnMNHV.txt » et « TxMNHV.txt » qui ont été obtenus en fin d'homogénéisation de la température.

L'analyse se fait ensuite station par station.

Pour la version semi-automatique :

(Utilisation du fichier de commande AnalyseOrgTnx.R)

Attention, il faut mettre un point d'arrêt (utilisation partielle) après chaque résumé de régression (`summary(fit)`) pour ajustement manuel des coefficients d'entrée (`start=list(...)`) en respect de la convention trigonométrique soit amplitude positive et phase comprise entre $-\Pi$ et $+\Pi$.

```
# Programme d'analyse de l'organisation chronologique d'une série de Tn et Tx (mono-station à base fixe)
#####

# Initialisations, chemin, nom du fichier des données Tn, Tx, numéro INSEE de la station

Chemin<-"~/home/debian/Documents/Rfiles/Organisation/"
NomFicTn="TnMNHV.txt" (à remplir manuellement)
NomFicTx<-"TxMNHV.txt" (à remplir manuellement)
NumStat<-"03060" (à remplir manuellement)

...

# Organisation de long terme
#####

dfTn<-data.frame(x=X,y=YTn)
dfTx<-data.frame(x=X,y=YTx)

# On commence par les Tn

fit<-nls(y ~ (C+TEND*x/36524
+ (AMP1C+AMPLTEND*(x+182)/36524)
* sin(2*acos(-1)*x/365.25+ (PHAS1C+PHAS1TEND*(x+182)/36524))
+AMP2*sin(2*acos(-1)*x/182.625+PHAS2)),
data=dfTn,
start=list(C=6, TEND=3, AMP1C=6.5, AMP1TEND=-1, PHAS1C=-2, PHAS1TEND=.3, AMP2=.5, PHAS2=.4))
summary(fit) (point d'arrêt!!)

...
```

```

# On continue par les Tx

fit<-nls(y ~ (C+TEND*x/36524
+(AMP1C+AMP1TEND*(x+182)/36524)
*sin(2*acos(-1)*x/365.25+(PHAS1C+PHAS1TEND*(x+182)/36524))
+AMP2*sin(2*acos(-1)*x/182.625+PHAS2)),
data=dfTx,
start=list(C=15,TEND=3,AMP1C=9,AMP1TEND=2,PHAS1C=-2,PHAS1TEND=.5,AMP2=1,PHAS2=-.4))
summary(fit) (point d'arrêt!!)

...

# Organisation de court terme
#####

dfTnSSorgO<-data.frame(x=X,y=Tn0,Tn1,Tn2,Tn3,Tn4,Tn5,Tn6,Tn7,Tx1,Tx2,Tx3,Tx4,Tx5,Tx6,Tx7)
dfTxSSorgO<-data.frame(x=X,y=Tx0,Tx1,Tx2,Tx3,Tx4,Tx5,Tx6,Tx7,Tn0,Tn1,Tn2,Tn3,Tn4,Tn5,Tn6,Tn7)

# Paramètre Tn

fit<-nls(y ~ ((A1C+A1TEND*x/36524+A1AMP1*sin(2*acos(-1)*x/365.25+A1PHAS1)+A1AMP2*sin(2*acos(-1)*x/
182.625+A1PHAS2))*Tn1
+(A2C+A2TEND*x/36524+A2AMP1*sin(2*acos(-1)*x/365.25+A2PHAS1)+A2AMP2*sin(2*acos(-1)*x/
182.625+A2PHAS2))*Tn2
+(A3C)*Tn3
+(A5C)*Tn5
+(A6C)*Tn6
+(B1C+B1TEND*x/36524+B1AMP1*sin(2*acos(-1)*x/365.25+B1PHAS1)+B1AMP2*sin(2*acos(-1)*x/
182.625+B1PHAS2))*Tx1
+(B2C+B2TEND*x/36524+B2AMP1*sin(2*acos(-1)*x/365.25+B2PHAS1)+B2AMP2*sin(2*acos(-1)*x/
182.625+B2PHAS2))*Tx2
+(B3C+B3TEND*x/36524+B3AMP1*sin(2*acos(-1)*x/365.25+B3PHAS1)+B3AMP2*sin(2*acos(-1)*x/
182.625+B3PHAS2))*Tx3
+(B6C)*Tx6
+(B7C)*Tx7),
data=dfTnSSorgO,
start=list(A1C=.60,A1TEND=-.7,A1AMP1=.09,A1PHAS1=1.6,A1AMP2=.06,A1PHAS2=-1.5,
A2C=-.1,A2TEND=.2,A2AMP1=.01,A2PHAS1=-1.3,A2AMP2=.04,A2PHAS2=.7,
A3C=.05,A5C=.01,A6C=.02,
B1C=.44,B1TEND=.07,B1AMP1=.02,B1PHAS1=-3,B1AMP2=.09,B1PHAS2=2.1,
B2C=-.1,B2TEND=-.02,B2AMP1=.01,B2PHAS1=1.5,B2AMP2=.01,B2PHAS2=1.7,
B3C=-.04,B3TEND=-.02,B3AMP1=.01,B3PHAS1=-1.4,B3AMP2=.02,B3PHAS2=-2.4,
B6C=-.02,B7C=-.01))

summary(fit) (point d'arrêt!!)

...

# Paramètre Tx

fit<-nls(y ~ ((A0C+A0TEND*x/36524+A0AMP1*sin(2*acos(-1)*x/365.25+A0PHAS1)+A0AMP2*sin(2*acos(-1)*x/
182.625+A0PHAS2))*Tn0
+(A1C+A1TEND*x/36524+A1AMP1*sin(2*acos(-1)*x/365.25+A1PHAS1)+A1AMP2*sin(2*acos(-1)*x/
182.625+A1PHAS2))*Tn1
+(A2AMP2*sin(2*acos(-1)*x/182.625+A2PHAS2))*Tn2
+(A3AMP2*sin(2*acos(-1)*x/182.625+A3PHAS2))*Tn3
+(B1C+B1TEND*x/36524+B1AMP1*sin(2*acos(-1)*x/365.25+B1PHAS1)+B1AMP2*sin(2*acos(-1)*x/
182.625+B1PHAS2))*Tx1
+(B2C+B2TEND*x/36524+B2AMP1*sin(2*acos(-1)*x/365.25+B2PHAS1)+B2AMP2*sin(2*acos(-1)*x/
182.625+B2PHAS2))*Tx2
+(B3C+B3AMP2*sin(2*acos(-1)*x/182.625+B3PHAS2))*Tx3
+(B4C)*Tx4
+(B6C)*Tx6),
data=dfTxSSorgO,
start=list(A0C=.5,A0TEND=.6,A0AMP1=.04,A0PHAS1=0,A0AMP2=.13,A0PHAS2=1.9,
A1C=-.12,A1TEND=.15,A1AMP1=.08,A1PHAS1=1.3,A1AMP2=.09,A1PHAS2=-2.7,
A2AMP2=.05,A2PHAS2=.5,
A3AMP2=.02,A3PHAS2=-2.4,
B1C=.66,B1TEND=.01,B1AMP1=.07,B1PHAS1=-1.8,B1AMP2=.09,B1PHAS2=-1.1,
B2C=-.02,B2TEND=-.1,B2AMP1=.02,B2PHAS1=2.5,B2AMP2=.04,B2PHAS2=1.4,
B3C=.04,B3AMP2=.03,B3PHAS2=-.08,
B4C=.02,B6C=.02))

summary(fit) (point d'arrêt!!)

...

# Calcul des écarts-type mensuels
#####
...

# Organisation de long terme

dfETn<-data.frame(x=XMois,y=ETMens[,2])
dfETx<-data.frame(x=XMois,y=ETMens[,3])

# Optimisation directe du schéma d'organisation de l'écart-type mensuel, paramètre Tn

fit<-nls(y ~ (C+TEND*x/1200
+AMP1*sin(2*acos(-1)*x/12+PHAS1)
+AMP2*sin(2*acos(-1)*x/6+PHAS2)),
data=dfETn,

```

```

      start=list (C=1.9,TEND=.4,AMP1=.30,PHAS1=1.2,AMP2=.07,PHAS2=2)
summary(fit) (point d'arrêt!!)
...
# Optimisation directe du schéma d'organisation de l'écart-type mensuel, paramètre Tx
fit<-nls(y ~ (C+TEND*x/1200
+AMP1*sin(2*acos(-1)*x/12+PHAS1)
+AMP2*sin(2*acos(-1)*x/6+PHAS2)),
data=dfETx,
start=list(C=2.2,TEND=-.14,AMP1=.50,PHAS1=-1.7,AMP2=.21,PHAS2=-1.5))
summary(fit) (point d'arrêt!!)
...
# Sorties impressions basiques
summary(TnxSortie)
sd(TnSaETu,na.rm=TRUE)
sd(TxSaETu,na.rm=TRUE)

```

Outre le non-respect de la convention trigonométrique, il peut arriver que le processus ne converge pas en raison du grand nombre de coefficients à estimer (organisation de court terme en particulier) auquel cas on pourra s'inspirer d'une station similaire pour confectionner une nouvelle liste de valeurs de démarrage.

En sortie on récupère les fichiers de coefficients-coordonnées d'organisation :

- « Tn|xSSorgOnnnnn.txt » pour les fichiers d'organisation de long terme
- « Tnx|xnSauLVOnnnnn.txt » pour les fichiers d'organisation de court terme
- « ETn|xSSorgOnnnnn.txt » pour les fichiers d'organisation de l'écart-type mensuel

On récupère aussi les fichiers de données observées, désorganisées progressivement :

- « Tnxnnnnn.txt » en données quotidiennes de base et « ETTnxnnnnn.txt » en écart-type mensuel

où nnnnn est le numéro INSEE (en 5 ou 6 chiffres) de la station considérée.

On sauvegardera dans des répertoires spécifiques à chaque station ces 8 fichiers, de même que les valeurs d'écart-type affichées à la fin.

Pour la version automatique :

(Utilisation du fichier de commande AnalyseOrgTnxAuto.R)

L'intérêt de cette version automatique est qu'elle ne nécessite *a priori* aucune intervention sauf l'inscription initiale des noms de fichiers d'observation et du numéro de la station. En effet, les diverses régressions sont effectuées 2 fois avec, en entrée de la seconde fois, les valeurs finales de la première exécution sur lesquelles la convention trigonométrique a été appliquée. De plus, l'affichage final de moyennes et écart-types est remplacé par un stockage de ces valeurs dans des fichiers, ce qui facilite un peu plus l'automatisation des projections.

Cependant, ceci ne supprime pas l'éventualité d'une non-convergence. Or, ce cas se produit en moyenne 1 fois sur 10. Il faut alors effectuer des exécutions alternées des versions automatiques et semi-automatiques (point d'arrêt par point d'arrêt) pour aboutir à des listes de valeurs de départ des régressions qui assurent la convergence.

```

# Programme d'analyse de l'organisation chronologique d'une série de Tn et Tx (mono-station à base fixe).
#####
# Version avec respect automatisé de la convention trigonométrique et stockage des moyennes et écart-types
# gaussiens des variables résiduelles.
#####
# Initialisations, chemin, nom du fichier des données Tn, Tx, numéro INSEE de la station

Chemin<-" /home/debian/Documents/Rfiles/Organisation/"
NomFicTn<-"TnMNHV.txt" (à remplir manuellement)
NomFicTx<-"TxMNHV.txt" (à remplir manuellement)
NumStat<-"03060" (à remplir manuellement)

```

```

...
# Organisation de long terme
#####

dfTn<-data.frame(x=X,y=YTn)
dfTx<-data.frame(x=X,y=YTx)

# On commence par les Tn
# Passage initial pour guess

fit<-nls(y ~ (C+TEND*x/36524
+(AMP1C+AMP1TEND*(x+182)/36524)
*sin(2*acos(-1)*x/365.25+(PHAS1C+PHAS1TEND*(x+182)/36524))
+AMP2*sin(2*acos(-1)*x/182.625+PHAS2)),
data=dfTn,
start=list(C=6,TEND=3,AMP1C=6.5,AMP1TEND=-1,PHAS1C=-2,PHAS1TEND=.3,AMP2=.5,PHAS2=.4))

#summary(fit)

# Guess pour passage 2, avec application de la convention trigonometrique
GuessC<-as.numeric(coef(fit)[1]) ; GuessTEND<-as.numeric(coef(fit)[2])
GuessAMP1C<-as.numeric(coef(fit)[3]) ; GuessAMP1TEND<-as.numeric(coef(fit)[4])
GuessPHAS1C<-as.numeric(coef(fit)[5]) ; GuessPHAS1TEND<-as.numeric(coef(fit)[6])
GuessAMP2<-as.numeric(coef(fit)[7]) ; GuessPHAS2<-as.numeric(coef(fit)[8])

if (GuessAMP1C < 0)
{GuessAMP1C<-(-GuessAMP1C) ; GuessAMP1TEND<-(-GuessAMP1TEND) ; GuessPHAS1C<-GuessPHAS1C+acos(-1)}
if (GuessPHAS1C < (-acos(-1)))
{Nb2Pi<-(-GuessPHAS1C+acos(-1)) %/% (2*acos(-1)) ; GuessPHAS1C<-GuessPHAS1C+Nb2Pi*2*acos(-1)}
if (GuessPHAS1C > (acos(-1)))
{Nb2Pi<-(-GuessPHAS1C+acos(-1)) %/% (2*acos(-1)) ; GuessPHAS1C<-GuessPHAS1C-Nb2Pi*2*acos(-1)}

if (GuessAMP2 < 0)
{GuessAMP2<-(-GuessAMP2) ; GuessPHAS2<-GuessPHAS2+acos(-1)}
if (GuessPHAS2 < (-acos(-1)))
{Nb2Pi<-(-GuessPHAS2+acos(-1)) %/% (2*acos(-1)) ; GuessPHAS2<-GuessPHAS2+Nb2Pi*2*acos(-1)}
if (GuessPHAS2 > (acos(-1)))
{Nb2Pi<-(-GuessPHAS2+acos(-1)) %/% (2*acos(-1)) ; GuessPHAS2<-GuessPHAS2-Nb2Pi*2*acos(-1)}

# Passage 2, pour vérification impression summary et stockage

fit<-nls(y ~ (C+TEND*x/36524
+(AMP1C+AMP1TEND*(x+182)/36524)
*sin(2*acos(-1)*x/365.25+(PHAS1C+PHAS1TEND*(x+182)/36524))
+AMP2*sin(2*acos(-1)*x/182.625+PHAS2)),
data=dfTn,
start=list(C=GuessC,TEND=GuessTEND,AMP1C=GuessAMP1C,AMP1TEND=GuessAMP1TEND,
PHAS1C=GuessPHAS1C,PHAS1TEND=GuessPHAS1TEND,AMP2=GuessAMP2,PHAS2=GuessPHAS2))

summary(fit)

...

# Sorties impressions basiques
summary(TnxSortie)
sd(TnSaETu,na.rm=TRUE);sd(TxSaETu,na.rm=TRUE)
sd(Tn0,na.rm=TRUE);sd(Tx0,na.rm=TRUE)

# Stockage de ces impressions pour reprise automatique dans procédure de projection
MoySDTnSaETu<-cbind(mean(TnSaETu,na.rm=TRUE),sd(TnSaETu,na.rm=TRUE))
MoySDTxSaETu<-cbind(mean(TxSaETu,na.rm=TRUE),sd(TxSaETu,na.rm=TRUE))
MoySDTnSSorgO<-cbind(mean(Tn0,na.rm=TRUE),sd(Tn0,na.rm=TRUE))
MoySDTxSSorgO<-cbind(mean(Tx0,na.rm=TRUE),sd(Tx0,na.rm=TRUE))

...

```

En sortie on récupère, en plus de la version semi-automatique les fichiers de moyennes et écart-types utilisés dans les projections, à stocker dans le même répertoire spécifique :

- « MoySDTn|xSSorgOnnnnn.txt » pour les résidus après extraction de l'organisation de long terme
- « MoySDTn|xSaETunnnnn.txt » pour les résidus finaux

4.3.2. ETP

Ce paramètre ayant été homogénéisé en une seule zone, sur la totalité du Massif central, on importera dans l'environnement R le fichier « ETPMfHV.txt », en plus des deux fichiers « TnMNHV.txt » et « TxMNHV.txt » puisque l'analyse et l'extraction des organisations chronologiques se fait conjointement.

Déroulé pour la version semi-automatique de l'analyse :

(Utilisation du fichier de commande AnalyseOrgETPTnx.R)

```
# Programme d'analyse de l'organisation chronologique d'une série d'ETP, Tn et Tx (mono-station à base fixe)
#####

# Initialisations, chemin, nom du fichier des données ETP, Tn, Tx, numéro INSEE de la station

Chemin<-"~/home/debian/Documents/Rfiles/Organisation/"
NomFicETP<-"ETPMfHV.txt" (à remplir manuellement)
NomFicTn<-"TnMNHV.txt" (à remplir manuellement)
NomFicTx<-"TxMNHV.txt" (à remplir manuellement)
NumStat<-"58062" (à remplir manuellement)

...

# Organisation de long terme
#####

dfETP<-data.frame(x=X,y=YETP)
dfTn<-data.frame(x=X,y=YTn)
dfTx<-data.frame(x=X,y=YTx)

# On commence par les Tn

...

# On continue par les Tx

...

# On termine par les ETP

fit<-nls(y ~ (C+TEND*x/36524
+ (AMP1C+AMP1TEND*(x+182)/36524)
* sin(2*acos(-1)*x/365.25+(PHAS1C+PHAS1TEND*(x+182)/36524))
+ AMP2*sin(2*acos(-1)*x/182.625+PHAS2)),
data=dfETP,
start=list(C=2, TEND=.7, AMP1C=1.6, AMP1TEND=1, PHAS1C=-1.6, PHAS1TEND=.2, AMP2=.3, PHAS2=.7))
summary(fit) (point d'arrêt!!)

...

# Organisation de court terme
#####

...

# Paramètre ETP

fit<-nls(y ~ ((A0C+A0TEND*x/36524+A0AMP1*sin(2*acos(-1)*x/365.25+A0PHAS1)+A0AMP2*sin(2*acos(-1)*x/
182.625+A0PHAS2))*Tn0
+ (A1C+A1TEND*x/36524+A1AMP1*sin(2*acos(-1)*x/365.25+A1PHAS1)+A1AMP2*sin(2*acos(-1)*x/
182.625+A1PHAS2))*Tn1
+ (A5AMP2*sin(2*acos(-1)*x/182.625+A5PHAS2))*Tn5
+ (A6AMP2*sin(2*acos(-1)*x/182.625+A6PHAS2))*Tn6
+ (B0C+B0TEND*x/36524+B0AMP1*sin(2*acos(-1)*x/365.25+B0PHAS1)+B0AMP2*sin(2*acos(-1)*x/
182.625+B0PHAS2))*Tx0
+ (B1C+B1TEND*x/36524+B1AMP1*sin(2*acos(-1)*x/365.25+B1PHAS1)+B1AMP2*sin(2*acos(-1)*x/
182.625+B1PHAS2))*Tx1
+ (B2C+B2TEND*x/36524+B2AMP2*sin(2*acos(-1)*x/182.625+B2PHAS2))*Tx2
+ (B3C+B3AMP1*sin(2*acos(-1)*x/365.25+B3PHAS1))*Tx3
+ (B7C+B7AMP1*sin(2*acos(-1)*x/365.25+B7PHAS1))*Tx7
+ (C1C+C1AMP1*sin(2*acos(-1)*x/365.25+C1PHAS1)+C1AMP2*sin(2*acos(-1)*x/182.625+C1PHAS2))*ETP1
+ (C2C+C2AMP1*sin(2*acos(-1)*x/365.25+C2PHAS1)+C2AMP2*sin(2*acos(-1)*x/182.625+C2PHAS2))*ETP2
+ (C3C)*ETP3
+ (C4C)*ETP4
+ (C5C)*ETP5
+ (C6C)*ETP6
+ (C7C+C7TEND*x/36524)*ETP7),
data=dfETPSSorg0,
start=list(A0C=-.05, A0TEND=.12, A0AMP1=.03, A0PHAS1=2.1, A0AMP2=.002, A0PHAS2=1.9,
A1C=-.02, A1TEND=.03, A1AMP1=.005, A1PHAS1=1.7, A1AMP2=.005, A1PHAS2=.1,
A5AMP2=.004, A5PHAS2=1.2,
A6AMP2=.004, A6PHAS2=-1,
B0C=.21, B0TEND=-.25, B0AMP1=.1, B0PHAS1=-1.5, B0AMP2=.008, B0PHAS2=0,
B1C=-.05, B1TEND=.07, B1AMP1=.04, B1PHAS1=1.2, B1AMP2=.006, B1PHAS2=-2.3,
B2C=-.01, B2TEND=.1, B2AMP2=.01, B2PHAS2=3,
B3C=-.005, B3AMP1=.005, B3PHAS1=2,
B7C=-.005, B7AMP1=.005, B7PHAS1=1.8,
C1C=.4, C1AMP1=.1, C1PHAS1=1.5, C1AMP2=.06, C1PHAS2=1.2,
C2C=.04, C2AMP1=.04, C2PHAS1=-1.5, C2AMP2=.04, C2PHAS2=1.5,
C3C=.03, C4C=.02, C5C=.02, C6C=.01,
C7C=.05, C7TEND=-.1))
summary(fit) (point d'arrêt!!)

...

# Calcul des écarts-type mensuels
#####

...
```



```
# Optimisation directe du schéma d'organisation de l'écart-type mensuel, paramètre ETP
fit<-nls(y ~ (C+TEND*x/1200
+ (AMP1C+AMP1TEND*(x+6)/1200)*sin(2*acos(-1)*x/12+PHAS1)
+AMP2*sin(2*acos(-1)*x/6+PHAS2)),
data=dfEETP,
start=list(C=.4,TEND=.15,AMP1C=.15,AMP1TEND=.2,PHAS1=-1.8,AMP2=.02,PHAS2=2.5),
algorithm="port",
upper=list(C=99,TEND=.28,AMP1C=99,AMP1TEND=99,PHAS1=99,AMP2=99,PHAS2=99))
summary(fit) (point d'arrêt!!)
...
# Sorties impressions basiques
summary(ETPTnxSortie)
sd(TnSaETu,na.rm=TRUE);sd(TxSaETu,na.rm=TRUE);sd(ETPSaETu,na.rm=TRUE)
sd(Tn0,na.rm=TRUE);sd(Tx0,na.rm=TRUE);sd(ETP0,na.rm=TRUE)
```

En sortie on récupère les fichiers de coefficients-coordonnées d'organisation :

- « Tn|xSSorgOnnnnn.txt » et « ETPSSorgOnnnnn.txt » pour les fichiers d'organisation de long terme
- « Tn|xSSauLVOnnnnn.txt » et « ETPSSauLVOnnnnn.txt » pour les fichiers d'organisation de court terme
- « ETn|xSSorgOnnnnn.txt » et « EETPSSorgOnnnnn.txt » pour les fichiers d'organisation de l'écart-type mensuel

On récupère aussi les fichiers de données observées, désorganisées progressivement :

- « ETPTnxnnnnn.txt » en données quotidiennes de base et « ETETPTnxnnnnn.txt » en écart-type mensuel

où nnnnn est le numéro INSEE (en 5 ou 6 chiffres) de la station considérée.

On sauvegardera dans des répertoires spécifiques à chaque station ces 11 fichiers, de même que les valeurs d'écart-type affichées à la fin.

Pour la version automatique :

(Utilisation du fichier de commande AnalyseOrgETPTnxAuto.R)

L'utilisation en est strictement équivalente à celle relative aux températures seules, donc on ne la rappellera pas.

En sortie on récupère, en plus de la version semi-automatique les fichiers de moyennes et écart-types utilisés dans les projections, à stocker dans le même répertoire spécifique :

- « MoySDTn|xSSorgOnnnnn.txt » et « MoySDETPSSorgOnnnnn.txt » pour les résidus après extraction de l'organisation de long terme
- « MoySDTn|xSaETunnnnn.txt » et « MoySDETPSaETunnnnn.txt » pour les résidus finaux

4.3.3. Précipitations

On prendra comme exemple le cas de la zone Nord du Massif central. Importer dans l'environnement R le fichier « RRMNHV.txt » en plus des deux fichiers « TnMNHV.txt » et « TxMNHV.txt » puisque l'analyse et l'extraction des organisations chronologiques de court terme se fait avec le support des séries complètes de température.

Comme le réseau de mesure de précipitations est plus fourni que celui des températures, on se retrouve dans deux cas de figures distincts. Soit il existe une mesure de températures sur la même station et on utilisera (sauf exception) les données locales de températures comme support de

l'extraction de l'organisation de court terme des précipitations. Soit la station ne mesure que les précipitations et il faut alors se rattacher à une station qui mesure effectivement les températures. Evidemment, ceci ne peut se faire au hasard ni même avec un critère de distance minimale. On étudiera donc distinctement ce cas où les températures sont déportées et on développera les critères de choix de la station thermique de référence.

4.3.3.1. Usage de températures locales

Déroulé pour la version semi-automatique de l'analyse :

(Utilisation du fichier de commande AnalyseOrgLRR.R)

```
# Programme d'analyse de l'organisation chronologique d'une série de précipitations LRR,
# via les correspondances avec Tn et Tx (températures locales)
#####

# Initialisations, chemin, nom du fichier des données RR, Tn, Tx, numéro INSEE de la station

Chemin<-"/home/debian/Documents/Rfiles/Organisation/"
NomFicRR<-"RRMNHV.txt" (à remplir manuellement)
NomFicTn<-"TnMNHV.txt" (à remplir manuellement)
NomFicTx<-"TxMNHV.txt" (à remplir manuellement)
NumStat<-"63003" (à remplir manuellement)

...

# Organisation de long terme
#####

# Organisation incluse dans celle de court-terme, donc préparation par création de variables décalées

...

# Organisation de court terme
#####

dfLRRSSorgO<-
data.frame(x=X,y=LRR0,LRR1,LRR2,LRR3,LRR4,LRR5,LRR6,LRR7,Tn0,Tn1,Tn2,Tn3,Tn4,Tn5,Tn6,Tn7,Tx0,Tx1,Tx2,Tx3,Tx4,Tx5,T
x6,Tx7)

# Paramètre LRR

fit<-nls(y ~ (C+TEND*x/36524+AMP1*sin(2*acos(-1)*x/365.25+PHAS1)+AMP2*sin(2*acos(-1)*x/182.625+PHAS2)
+(A0C+A0TEND*x/36524+A0AMP1*sin(2*acos(-1)*x/365.25+A0PHAS1)+A0AMP2*sin(2*acos(-1)*x/
182.625+A0PHAS2))*Tn0
+(A1C)*Tn1
+(A3AMP2*sin(2*acos(-1)*x/182.625+A3PHAS2))*Tn3
+(B0C+B0TEND*x/36524+B0AMP1*sin(2*acos(-1)*x/365.25+B0PHAS1)+B0AMP2*sin(2*acos(-1)*x/
182.625+B0PHAS2))*Tx0
+(B1C+B1AMP1*sin(2*acos(-1)*x/365.25+B1PHAS1)+B1AMP2*sin(2*acos(-1)*x/182.625+B1PHAS2))*Tx1
+(B2C+B2AMP1*sin(2*acos(-1)*x/365.25+B2PHAS1))*Tx2
+(B4AMP2*sin(2*acos(-1)*x/182.625+B4PHAS2))*Tx4
+(C1C+C1TEND*x/36524+C1AMP1*sin(2*acos(-1)*x/365.25+C1PHAS1)+C1AMP2*sin(2*acos(-1)*x/
182.625+C1PHAS2))*LRR1
+(C2C+C2AMP1*sin(2*acos(-1)*x/365.25+C2PHAS1))*LRR2
+(C3C)*LRR3
+(C6C+C6AMP1*sin(2*acos(-1)*x/365.25+C6PHAS1))*LRR6,
data=dfLRRSSorgO,
start=list(C=-.03,TEND=-.12,AMP1=.13,PHAS1=-.08,AMP2=.1,PHAS2=2.8,
A0C=.04,A0TEND=.02,A0AMP1=.005,A0PHAS1=-2.9,A0AMP2=.01,A0PHAS2=-.7,
A1C=-.006,
A3AMP2=.003,A3PHAS2=-2.1,
B0C=-.04,B0TEND=.001,B0AMP1=.03,B0PHAS1=1.7,B0AMP2=.02,B0PHAS2=1.5,
B1C=.03,B1AMP1=.03,B1PHAS1=-1.6,B1AMP2=.01,B1PHAS2=-1.5,
B2C=-.01,B2AMP1=.003,B2PHAS1=.9,
B4AMP2=.003,B4PHAS2=-1.7,
C1C=.25,C1TEND=-.29,C1AMP1=.09,C1PHAS1=1.2,C1AMP2=.03,C1PHAS2=3.0,
C2C=.06,C2AMP1=.03,C2PHAS1=-1,
C3C=.02,
C6C=.02,C6AMP1=.01,C6PHAS1=-.8)

summary(fit) (point d'arrêt!!)

...

# Calcul des écarts-type mensuels
#####

...

dfELRR<-data.frame(x=XMois,y=ETMens[,2])

# Optimisation directe du schéma d'organisation de l'écart-type mensuel, paramètre LRR
```

```

fit<-nls(y ~ (C+TEND*x/1200
+AMP1*sin(2*acos(-1)*x/12+PHAS1)
+AMP2*sin(2*acos(-1)*x/6+PHAS2)),
data=dfELRR,
start=list(C=.55,TEND=-.13,AMP1=.03,PHAS1=-2.8,AMP2=.02,PHAS2=2.8))
summary(fit) (point d'arrêt!!)

...

# Sorties impressions basiques
summary(LRRSortie)
sd(LRRSaETu,na.rm=TRUE)
sd(LRR0,na.rm=TRUE)

```

En sortie on récupère les fichiers de coefficients-coordonnées d'organisation :

- « LRRSSauLVOnnnnn.txt » pour le fichier d'organisation de court terme
- « ELRRSSorgOnnnnn.txt » pour le fichier d'organisation de l'écart-type mensuel

On récupère aussi les fichiers de données observées, désorganisées progressivement :

- « LRRnnnnn.txt » en données quotidiennes de base et « ETLRRnnnnn.txt » en écart-type mensuel

où nnnnn est le numéro INSEE (en 5 ou 6 chiffres) de la station considérée.

Sauvegarder dans des répertoires spécifiques à chaque station ces 4 fichiers, de même que les valeurs d'écart-type affichées à la fin.

Pour la version automatique :

(Utilisation du fichier de commande AnalyseOrgLRRAuto.R)

L'utilisation en est strictement équivalente à celle relative aux températures seules, donc on ne la rappellera pas.

En sortie, on récupère, en plus de la version semi-automatique, les fichiers de moyennes et écart-types utilisés dans les projections. Ces fichiers sont à stocker dans le même répertoire spécifique :

- « MoySDLRRnnnnn.txt » pour la série complète, soit après l'extraction inexistante de l'organisation de long terme
- « MoySDLRRSaETunnnnn.txt » pour les résidus finaux

4.3.3.2. Usage de températures déportées

Avant de pouvoir utiliser les fichiers de calcul et d'extraction des organisations chronologiques, il faut d'abord faire le choix de la station thermique de référence. Une procédure quasi-automatisée a été conçue à cet effet. Elle effectue, pour l'ensemble d'une zone, une analyse des corrélations entre les séries de précipitations transformées (LRR) et les séries de Tn et Tx. Pour éviter les risques de non-convergence, ce qui obligerait à traiter longuement cet aspect non-principal, ce n'est pas un schéma aussi sophistiqué que celui de la détermination des inter-organisations de court terme qui a été utilisé mais une régression saisonnalisée (coefficients sur Tn et Tx variables en tendance et en cycles de 6 et 12 mois fixes).

(Utilisation du fichier de commande SelectTnxDistOrgLRR.R)

```

# Programme de sélection semi-automatique des stations déportées de Tn et Tx pour
# l'analyse de l'organisation chronologique d'une série de précipitations LRR
#####

```

```

# Initialisations, chemin, nom des fichiers de données RR, Tn, Tx
# List des numéros INSEE de la station RR
# Liste des numéros des stations TnTx

Chemin<="/home/debian/Documents/Rfiles/Organisation/" (à remplir manuellement)
NomFicRR<-"RRMSHV.txt" (à remplir manuellement)
NomFicTn<-"TnMSHV.txt" (à remplir manuellement)
NomFicTx<-"TxMSHV.txt" (à remplir manuellement)
ListStatRR<-c("11081","12036","19016","19019","19094","19100","19129","19133","19139","19157",
"19206","19226","19227","19258","19264","192882","34228","34233","46031","46038",
"46128","81163","82148") (toute la liste à remplir manuellement)
NbStatRR=NROW(ListStatRR)
ListStatT<-c("12145","12214","12254","12300","34046","46005","46083","46127","46160","46197",
"81124","81182","19010","19029","19031","19033","19034","19036","19121","19125",
"19134","19136","19146","19148","19164","19237","19272","192751","192754","192884")
NbStatT=NROW(ListStatT) (toute la liste à remplir manuellement)
ListSDRes<-numeric(NbStatT)

# On importe les fichier de précipitations et de températures
NomFichier<-paste(Chemin,NomFicRR,sep="")
DonneesRR <-
  read.table(NomFichier,
  header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
NomFichier<-paste(Chemin,NomFicTn,sep="")
DonneesTn <-
  read.table(NomFichier,
  header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)
NomFichier<-paste(Chemin,NomFicTx,sep="")
DonneesTx <-
  read.table(NomFichier,
  header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE)

# Création vecteur temps en numéro d'observation
X<-seq(1,nrow(DonneesRR))

# Calcul de corrélation poste par poste de température
#####
for (IndicRR in seq(1,NbStatRR,1))
{
print(paste("Station RR : ",ListStatRR[IndicRR],sep=""));print(" ")
# Extraction de colonne relative à la station RR
NumCol<-0
for (i in seq(1,ncol(DonneesRR),1))
{if (substr(colnames(DonneesRR)[i],3,2+nchar(ListStatRR[IndicRR]))==ListStatRR[IndicRR])
{NumCol<-i}
}
YLRR<-log10(DonneesRR[,NumCol]+0.01+runif(nrow(DonneesRR),min=0,max=0.2))

for (IndicStat in seq(1,NbStatT,1))
{

# Extraction des colonnes relatives à la station T
NumCol<-0
for (i in seq(1,ncol(DonneesTn),1))
{if (substr(colnames(DonneesTn)[i],3,2+nchar(ListStatT[IndicStat]))==ListStatT[IndicStat])
{NumCol<-i}
}
YTn<-DonneesTn[,NumCol]

NumCol<-0
for (i in seq(1,ncol(DonneesTx),1))
{if (substr(colnames(DonneesTx)[i],3,2+nchar(ListStatT[IndicStat]))==ListStatT[IndicStat])
{NumCol<-i}
}
YTx<-DonneesTx[,NumCol]

dfLRR<-data.frame(x=X,y=YLRR,YTn,YTx)

# Regression saisonnière d'ordre 0

fit<-nls(y ~ ((A0C+A0TEND*x/36524+A0AMP1*sin(2*acos(-1)*x/365.25+A0PHAS1)+A0AMP2*sin(2*acos(-1)*x/
182.625+A0PHAS2))*YTn
+(B0C+B0TEND*x/36524+B0AMP1*sin(2*acos(-1)*x/365.25+B0PHAS1)+B0AMP2*sin(2*acos(-1)*x/
182.625+B0PHAS2))*YTx
+Const),
data=dfLRR,
start=list(A0C=.1,A0TEND=0,A0AMP1=.01,A0PHAS1=-2.3,A0AMP2=.01,A0PHAS2=-1.4,
B0C=-.4,B0TEND=-.01,B0AMP1=.01,B0PHAS1=.6,B0AMP2=.01,B0PHAS2=2.3,
Const=0))
ListSDRes[IndicStat]<-sd(resid(fit))
# print(paste("Station T : ",ListStatT[IndicStat]," SD résidus : ",as.character(sd(resid(fit))),sep=""))
}

# Impression ordonnée des valeurs d'écart-types des résidus et stations associées
for (i in seq(1,NbStatT,1))
{print(paste("Station T : ",ListStatT[order(ListSDRes)[i]]," SD résidus :
",as.character(ListSDRes[order(ListSDRes)[i]]),sep=""))}
}

```

A l'issue de cette procédure, on récupère un paquet d'impressions (une par station RR) avec la liste hiérarchisée des stations thermiques en corrélation décroissante. La méthode basique est de choisir la station qui apparaît à la première place.

Cependant, il est plutôt conseillé de reporter sur un tableau manuel la liste des 5 ou 10 meilleures stations thermiques en face de chaque station pluviométrique. Ceci permet de repérer une éventuelle sur-représentation d'une seule station thermique. En ce cas, on pourrait faire le choix de stations alternatives presque optimales, pour inclure une dose de stochasticité. Ceci résout aussi le cas de certains postes thermiques (particulièrement ceux situés en sommet de colline) qui se révèlent de médiocres prédicteurs des précipitations sur le schéma d'organisation complet.

C'est pour ce genre de raisons qu'il peut être préférable de choisir le cheminement relatif aux postes avec températures déportées, y compris pour des stations qui mesurent les températures localement. Dans notre exemple, les 2 stations de Bénévent (23021) et Guéret-Grancher (230967) ont subi ce traitement particulier.

Les postes thermiques de référence ayant été choisis, on peut passer à l'extraction des organisations chronologiques.

Déroulé pour la version semi-automatique de l'analyse :

(Utilisation du fichier de commande AnalyseOrgLRRTx.R)

A part le changement de nom de la procédure, la seule modification par rapport à la procédure avec températures locales est l'écriture manuelle supplémentaire du numéro INSEE de la station de référence thermique.

```
# Programme d'analyse de l'organisation chronologique d'une série de précipitations LRR,
# via les correspondances avec Tn et Tx (températures déportées)
#####
# Initialisations, chemin, nom du fichier des données RR, Tn, Tx, numéro INSEE de la station RR
# Numéro INSEE de la station T

Chemin<-"/home/debian/Documents/Rfiles/Organisation/" (à remplir manuellement)
NomFicRR<-"RRMNHV.txt" (à remplir manuellement)
NomFicTn<-"TnMNHV.txt" (à remplir manuellement)
NomFicTx<-"TxMNHV.txt" (à remplir manuellement)
NumStatRR<-"03161" (à remplir manuellement)
NumStatT<-"23008" (à remplir manuellement)

...
```

Les sorties sont identiques aux températures mesurées localement.

Pour la version automatique :

(Utilisation du fichier de commande AnalyseOrgLRRTxAuto.R)

L'utilisation en est strictement équivalente à celle relative aux températures mesurées localement, donc on ne la rappellera pas.

Une fois terminée l'analyse de l'organisation chronologique, on peut aborder l'étape finale du processus, à savoir la confection de projections climatiques proprement dites.

5. Projections climatiques

Durant le précédent chapitre, nous avons « dé-tricoté » les séries chronologiques de paramètres météorologiques en séparant ce qui était organisé, de long terme puis de court terme, d'un résidu dont l'étude « AP3C-Creuse » avait pu montrer que la variance diminuait progressivement, de même que son caractère aléatoire augmentait progressivement.

Pour les projections climatiques, d'une certaine manière, nous ferons le contraire. Cependant, le fait que la séparation entre évolution organisée et partie aléatoire ne soit ni unique ni parfaite, implique que le « re-tricotage » imposera un certain nombre de contrôles et de corrections pour être cohérent. Logiquement, ces contrôles et corrections sont plus importants encore pour les précipitations qui sont moins bien organisées chronologiquement et moins bien séparées.

5.1. Déroulement générique

Le fait d'être arrivé à un résidu (presque) aléatoire ouvre la possibilité de rejouer ce résidu par l'intermédiaire d'un générateur de nombre aléatoire. La contrainte minimale, pour que le résidu rejoué soit vraisemblable est qu'il respecte peu ou prou la distribution du résidu observé. De nombreuses possibilités ont été testées lors du projet « AP3C-Creuse », allant jusqu'à une combinaison linéaire de distributions classiques, en modulation d'une distribution normale généralisée. Ceci avait abouti à un ajustement très précis des 4 premiers « moments » de la distribution ou, de manière équivalente, de sa moyenne, de son écart-type, de sa dissymétrie et de son aplatissement.

Cependant, nous avons pu montrer que, par rapport à une simple gaussienne (qui est définie entièrement par sa moyenne et son écart-type), le gain sur la conformité de la variable météorologique complète (températures, précipitations...) reconstituée par rapport à la distribution observée, de la même variable complète, était négligeable. Ceci s'entend si on en reste à des distributions statiques dans le temps.

De plus, la recherche des combinaisons de distributions pour ajustement des dissymétrie et aplatissement est assez laborieuse. Dans la mesure où un des buts de AP3C est d'aboutir à une simplification et à une automatisation, il a été décidé d'en rester à une distribution gaussienne.

Pour décrire un large panel de possibilités, permettant éventuellement d'étudier des événements rares ou/et d'assurer une bonne stabilité statistique, il a été décidé de générer 10 000 séries aléatoires, aboutissant *in fine* à 10 000 projections climatiques.

Avec l'aide de quelques valeurs initiales vraisemblables, le schéma d'organisation de court terme est ensuite appliqué, pas à pas, depuis la date origine des observations. En ce sens, ce ne sont pas seulement des projections qui sont produites mais aussi (et surtout!) des réanalyses statistiques sur la période observée. On rappelle que ce schéma est multi-paramétrique. Le schéma de long terme (quand il y en a un) s'y combine de manière beaucoup plus simple, par une addition effectuée globalement.

Si la séparation entre organisations et caractère aléatoire était parfaite, on pourrait se contenter d'appliquer cette description assez simple. Mais ce n'est pas le cas et la distribution des valeurs réanalysées complètes diffère un peu (T_x), beaucoup (T_n , ETP) ou même énormément (RR) de la distribution observée correspondante. Il faut donc au minimum 2 itérations du processus décrit.

La première itération est faite sur un nombre réduit de simulations (100 projections) ce qui permet d'avoir néanmoins des valeurs stables pour la moyenne des distributions simulées, sur la

période d'observation. Ceci permet donc de calibrer une courbe de correction quantile-quantile vis à vis des observations (plus de détails en pages 147-153 du rapport technique de AP3C-Creuse). Lors de la seconde itération (en 10 000 exemplaires), cette courbe de correction est appliquée à l'ensemble des simulations.

Les spécificités de ce déroulement paramètre par paramètre seront vues dans des paragraphes ad-hoc. A noter que l'intégralité des procédures figure dans l'annexe « Fichiers Projections Climatiques.ods ».

5.2. La question du futur

Nous avons certes obtenu, par le processus décrit ci-avant, des réanalyses climatiques cohérentes avec les observations, à la fois en distribution et en organisations chronologiques. Mais qu'en est-il du futur ?

La méthode en elle-même n'en dit rien ou presque. Il faut aller chercher des considérations dans les projections climatiques physiques globales (les modèles GCM utilisés entre autres dans les rapports du GIEC) et les utiliser de manière relative à la trajectoire déjà engagée. On retrouvera ces éléments en téléchargement, dans la fiche méthode climat de AP3C (<https://www.sidam-massifcentral.fr/wp-content/uploads/2016/09/METHODE-CLIMATIQUE-2.pdf>).

Il faut également ajouter un élément supplémentaire, non publié mais crucial, qui a été obtenu en partie avec les résultats de l'étude qui a reçu le prix André Prud'homme en 1992 (article résumé sur <http://documents.irevues.inist.fr/handle/2042/53329>), à savoir que pour les organisations chronologiques qui ont une dérive tendancielle significative, cette dérive est proportionnelle, à l'ordre 1, à la tendance générale de la série.

Donnons un exemple, si la tendance générale de la température était de $-1,5^{\circ}\text{C}/\text{siècle}$ entre 1950 et 1980 et de $+4,5^{\circ}\text{C}/\text{siècle}$ à partir de 1980 alors le retardement du pic thermique annuel de 7 jours/siècle entre 1950 et 1980 se transforme en un avancement de 21 jours par siècle à partir de 1980. Le facteur de transformation est de -3 dans les 2 cas.

De sorte que les raisonnements qui sont produits sur les évolutions de tendance générale s'appliquent ipso-facto sur l'ensemble des organisations chronologiques.

En quelques mots, ce qui utilisé vers le futur (un futur climatique suffisamment proche pour être commun à tous les scénarios d'émission de GES) est la poursuite sans accélération des caractéristiques d'évolution observées. Et ceci constitue hélas un minorant très probable de l'évolution qui se produira réellement.

On va désormais aborder le détail des processus de projection, d'abord pour les températures et l'ETP, qui sont gérées de la même manière, puis pour les précipitations.

5.3. Températures et ETP

Les procédures correspondantes sont quasi-intégralement automatisées. Les seuls facteurs éventuellement délicats à gérer sont le temps de calcul qui se chiffre en heures et la mémoire vive nécessaire (jusque 30Go pour les ETP).

La première étape est le transfert vers l'espace de travail « R » des fichiers de configuration obtenus lors de l'analyse d'organisation, soit 12 fichiers pour les températures et 17 pour les ETP (qui incluent aussi les températures).

(Utilisation du fichier de commande SimTxxxxx.R ou SimETPTxxxxx.R)

On choisit un exemple de fichier concernant uniquement les températures. On déroule d'abord les 4 étapes du remontage des organisations : les résidus de variance uniforme (gaussiens),

les résidus avec variance organisée (ajout de l'organisation de la variance), les résidus autocorrélés (avec ajout de l'organisation de court terme) et enfin les valeurs simulées complètes (ajout de l'organisation de long terme).

```
# Simulation de Boen sur Lignon (Tnx 42019) (inscription manuelle à volonté)
# sur période 01/01/1980 - 31/12/2050, soit 25933 jours consécutifs.
# Ecriture fichiers sur période 01/01/1980 - 31/12/2050, pour statistiques.
# Ecriture fichier public sur 01/01/1980 - 31/12/2050 arrondi à 2 décimales
# pour limite de taille sous systèmes 32 bits.

# Récupération des coefficients d'organisation par fichiers. Résidus finaux gaussiens.
# Initialisation par fichiers des caractéristiques (moy,sd) des résidus et des premiers pas
# de la reconstruction de court-terme.
# Correction QQ post-simulation sur liste de valeurs à indices fixes, calcul parallélisé.

# Traitement des valeurs manquantes, facultatif si déjà inclus à l'installation de R
options(na.action = "na.exclude")

# Initialisations

Chemin<-"D:/Rfiles/SimulationTnx/R1/" (mettre manuellement le répertoire de travail, les sous-répertoires
R1,R2,...Ri sont utiles pour des exécutions multiples et parallèles de plusieurs stations différentes)
NumStat<-"42019" (mettre manuellement le numéro INSEE de la station)

SauvTxt<-TRUE (sortie optionnelle de fichiers intermédiaires, étape par étape de construction)
DebSauv<-19800101 (mettre le début des observations au format AAAAMMJJ)
DebSauvPub<-19800101 (idem)
NbPas<-25933 (nombre de jours entre le début des observations et la fin des projections)
NbSimul<-100 (c'est un minimum, mais on pourrait avoir le même nombre que pour la seconde exécution)

...

# On traite le cas d'un résidu final gaussien

# Création des tableaux de résidus avec variance uniforme
#####

# Pour Tn, on écrit le cas d'un résidu gaussien avec expression directe
# récupération des moyennes et écart-types précalculés
nomfichier<-paste(Chemin,"MoySDTnSaETu",NumStat,".txt",sep="")
MoySDTn<-
read.table(file=nomfichier,header=TRUE,sep=" ",dec=".")
MoyenneTn<-MoySDTn$MOY;ETypeTn<-MoySDTn$SD

SnSaETu <- matrix(rnorm(NbPas*NbSimul, mean=MoyenneTn, sd=ETypeTn),ncol=NbSimul)
colnames(SnSaETu) <- paste("Sn", 1:NbSimul, sep="")

# Ecriture format texte (optionnelle)
if(SauvTxt)
{nomfichier<-paste(Chemin,"SnSaETu.txt",sep="")
write.table(subset(cbind(DateJ,SnSaETu),(DateJ>=DebSauv)),
file=nomfichier, sep=" ",
col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA")}

# Ecriture format R Data, compact
SnSaETuRda<-cbind(DateJ,SnSaETu)
nomfichier<-paste(Chemin,"SnSaETu.Rda",sep="")
save(SnSaETuRda,file=nomfichier)
rm(SnSaETuRda)

...

# Création des résidus avec variance dé-uniformisée
#####

# Récupération des coefficients d'organisation de la variance

...

# Remplissage des tableaux
for(j in seq(1,NbSimul,1))
{for(i in seq(1,NbPas,1))
{SxnSauLVO[i,j]<-(SnSaETu[i,j])/TnC*(TnC+TnTend*(i+15)/36524
+TnAmp1*sin(2*acos(-1)*(i+15)/365.25+TnPhas1)
+TnAmp2*sin(2*acos(-1)*(i+15)/182.625+TnPhas2))

SxnSauLVO[i,j]<-(SxSaETu[i,j])/TxC*(TxC+TxTend*(i+15)/36524
+TxAmp1*sin(2*acos(-1)*(i+15)/365.25+TxPhas1)
+TxAmp2*sin(2*acos(-1)*(i+15)/182.625+TxPhas2))
}
}

...

# Création des résidus auto-corrélés, ajout du schéma d'autocorrélation
#####

# Récupération des coefficients d'organisation de court terme
```



```

...

# Remplissage des premiers éléments (non autocorrélés) des tableaux
# récupération des moyennes et écart-types précalculés
nomfichier<-paste(Chemin,"MoySDTnSSorgO",NumStat,".txt",sep="")
MoySDTn<-
read.table(file=nomfichier,header=TRUE,sep=" ",dec=".")
MoyenneTn<-MoySDTn$MOY;ETypeTn<-MoySDTn$SD
nomfichier<-paste(Chemin,"MoySDTxSSorgO",NumStat,".txt",sep="")
MoySDTx<-
read.table(file=nomfichier,header=TRUE,sep=" ",dec=".")
MoyenneTx<-MoySDTx$MOY;ETypeTx<-MoySDTx$SD

for(j in seq(1,NbSimul,1))
  {for(i in seq(1,7,1))
    {SnSSorgO[i,j]<-rnorm(1,mean=MoyenneTn,sd=ETypeTn)
     SxSSorgO[i,j]<-rnorm(1,mean=MoyenneTx,sd=ETypeTx)
    }
  }

# Remplissage des éléments suivants
for(i in seq(8,NbPas,1))
  {SnSSorgO[i,j]<-(SnSSorgO[i-1,j]*(TnA1C+TnA1Tend*i/36524+TnA1Amp1*sin(2*acos(-1)*i/
365.25+TnA1Phas1)+TnA1Amp2*sin(2*acos(-1)*i/182.625+TnA1Phas2))
  +SnSSorgO[i-2,j]*(TnA2C+TnA2Tend*i/36524+TnA2Amp1*sin(2*acos(-1)*i/
365.25+TnA2Phas1)+TnA2Amp2*sin(2*acos(-1)*i/182.625+TnA2Phas2))
  +SnSSorgO[i-3,j]*(TnA3C)
  +SnSSorgO[i-5,j]*(TnA5C)
  +SnSSorgO[i-6,j]*(TnA6C)
  +SxSSorgO[i-1,j]*(TnB1C+TnB1Tend*i/36524+TnB1Amp1*sin(2*acos(-1)*i/
365.25+TnB1Phas1)+TnB1Amp2*sin(2*acos(-1)*i/182.625+TnB1Phas2))
  +SxSSorgO[i-2,j]*(TnB2C+TnB2Tend*i/36524+TnB2Amp1*sin(2*acos(-1)*i/
365.25+TnB2Phas1)+TnB2Amp2*sin(2*acos(-1)*i/182.625+TnB2Phas2))
  +SxSSorgO[i-3,j]*(TnB3C+TnB3Tend*i/36524+TnB3Amp1*sin(2*acos(-1)*i/
365.25+TnB3Phas1)+TnB3Amp2*sin(2*acos(-1)*i/182.625+TnB3Phas2))
  +SxSSorgO[i-6,j]*(TnB6C)
  +SxSSorgO[i-7,j]*(TnB7C)
  +SxnSauLVO[i,j])
  SxSSorgO[i,j]<-(SnSSorgO[i,j]*(TxA0C+TxA0Tend*i/36524+TxA0Amp1*sin(2*acos(-1)*i/
365.25+TxA0Phas1)+TxA0Amp2*sin(2*acos(-1)*i/182.625+TxA0Phas2))
  +SnSSorgO[i-1,j]*(TxA1C+TxA1Tend*i/36524+TxA1Amp1*sin(2*acos(-1)*i/
365.25+TxA1Phas1)+TxA1Amp2*sin(2*acos(-1)*i/182.625+TxA1Phas2))
  +SnSSorgO[i-2,j]*(TxA2Amp2*sin(2*acos(-1)*i/182.625+TxA2Phas2))
  +SnSSorgO[i-3,j]*(TxA3Amp2*sin(2*acos(-1)*i/182.625+TxA3Phas2))
  +SxSSorgO[i-1,j]*(TxB1C+TxB1Tend*i/36524+TxB1Amp1*sin(2*acos(-1)*i/
365.25+TxB1Phas1)+TxB1Amp2*sin(2*acos(-1)*i/182.625+TxB1Phas2))
  +SxSSorgO[i-2,j]*(TxB2C+TxB2Tend*i/36524+TxB2Amp1*sin(2*acos(-1)*i/
365.25+TxB2Phas1)+TxB2Amp2*sin(2*acos(-1)*i/182.625+TxB2Phas2))
  +SxSSorgO[i-3,j]*(TxB3C+TxB3Amp2*sin(2*acos(-1)*i/182.625+TxB3Phas2))
  +SxSSorgO[i-4,j]*(TxB4C)
  +SxSSorgO[i-6,j]*(TxB6C)
  +SxnSauLVO[i,j])
  }
}

...

# Création des températures simulées, ajout de l'organisation de long terme
#####

# Récupération des coefficients d'organisation de long terme

...

# Remplissage du tableau
for(j in seq(1,NbSimul,1))
  {for(i in seq(1,NbPas,1))
    {Sn[i,j]<-(SnSSorgO[i,j]+TnC+TnTend*i/36524
  + (TnAmp1C+TnAmp1Tend*(i+182)/36524)*sin(2*acos(-1)*i/365.25+(TnPhas1C+TnPhas1Tend*(i+182)/36524))
  +TnAmp2*sin(2*acos(-1)*i/182.625+TnPhas2))
    Sx[i,j]<-(SxSSorgO[i,j]+TnC+TxTend*i/36524
  + (TxAmp1C+TxAmp1Tend*(i+182)/36524)*sin(2*acos(-1)*i/365.25+(TxPhas1C+TxPhas1Tend*(i+182)/36524))
  +TxAmp2*sin(2*acos(-1)*i/182.625+TxPhas2))
    }
  }

...

```

On déroule ensuite les opérations permettant d'effectuer la correction quantile-quantile post-simulation, soit le calcul des points pivots, une application de cette correction guidée par les points pivots sur un type de fichiers (ordonnés) permettant la vérification de la qualité de la correction, une autre application qui simule exactement ce qui sera fait lors du second passage de la procédure de projection (qui sera donc avec correction quantile-quantile), ce qui permet enfin des comparaisons de statistiques climatologiques basiques. On compare ce qui est obtenu directement sur les données observées (1980-2015 dans notre cas) avec ce qui est obtenu sur une année centrale des simulations correspondantes (1998) et en moyenne sur les simulations de la période observée (1980-2015).

```
#####
# Ensemble de procédures de préparation à la projection des températures avec correction QQ
#####
(Ce qui suit est donc spécifique au premier passage de la procédure de projection climatique)

# Initialisations : date de début, date de fin des obs
# liste des numéros de points ordonnés pivots pour correction QQ

DebObs<-19800101;FinObs<-20151231 (attention, il faut écrire ces dates manuellement)
AnDeb<-DebObs %/% 10000;AnFin<-FinObs %/% 10000
IndicFibo<-c(2,8,10,18,28,46,74,120,194,314,508,822,1330,2152,3482,5634,7516,9668,10998,11820,
12328,12642,12836,12956,13030,13076,13104,13122,13132,13140,13142,13148)
NbPts<-NROW(IndicFibo)
(ces points pivots ont été choisis en adoptant une progression de Fibonacci à partir des 2 premiers éléments. Ce choix a été fait par la lecture de nombreux graphes quantiles-quantiles. Un autre choix plus régulier peut être fait, mais il est préférable de conserver le même pour tout le réseau)

# On commence par la fusion des simulations avec l'observation, sur la période observée
#####
...

# On poursuit par le tri des valeurs colonne par colonne
#####
...

# Création des fichiers de points pivots pour correction QQ
#####
...

# Applications de la correction QQ sur échantillon de projections (pour contrôle)
#####
# PARTIE TRAITEMENT DES FICHIERS ORDONNES #
#####

# On s'occupe d'abord de Tn
#####
...

# Calcul des paramètres des segments de correction
# Importation des coordonnées des points

nomfichier<-paste(Chemin,"PtsQQTn.txt",sep="")
PtsQQTn <- as.matrix(
  read.table(file=nomfichier,header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE))

NbPts<-nrow(PtsQQTn)
ASn<-matrix(0,ncol=1,nrow=NbPts)
BSn<-matrix(0,ncol=1,nrow=NbPts)

for (i in seq(2,NbPts,1))
  {ASn[i]<-(PtsQQTn[i,3]-PtsQQTn[i-1,3])/(PtsQQTn[i,2]-PtsQQTn[i-1,2])
  BSn[i]<-PtsQQTn[i-1,3]-ASn[i]*PtsQQTn[i-1,2]
  }

# Application de la formule par segments
(si on veut que la formule soit appliquée point par point, ce qui est classique, il suffit de remplacer la liste d'indices de Fibonacci par la liste totale des indices)

for (i in seq(1,nrow(STnOrd),1))
  {if (SnMoy[i,1]<PtsQQTn[2,2])
    {SnMoyQQ[i,1]<-ASn[2]*SnMoy[i,1]+BSn[2]}
  for (j in seq(2,NbPts-2,1))
    {if (SnMoy[i,1]>=PtsQQTn[j,2] & SnMoy[i,1]<PtsQQTn[j+1,2])
      {SnMoyQQ[i,1]<-ASn[j+1]*SnMoy[i,1]+BSn[j+1]}
    }
  if (SnMoy[i,1]>=PtsQQTn[NbPts-1,2])
    {SnMoyQQ[i,1]<-ASn[NbPts]*SnMoy[i,1]+BSn[NbPts]}
  }

...

# Impression de contrôle des trente résidus de correction QQ les plus élevés
...

# On s'occupe ensuite de Tx
#####
...

# PARTIE TRAITEMENT DES FICHIERS NON-ORDONNES #
#####
(c'est celle qui sera appliquée lors du second passage)

# On s'occupe d'abord de Tn
#####
```

```

...
# Application de la formule par segments (calcul parallélisé)
(c'est essentiellement la même formule que précédemment)
for (k in seq(1,NbSimul,1))
  {SnQQ[,k]<-SnQQ[,k]+(as.numeric(Sn[,k]<PtsQQTn[2,2]))*(ASn[2]*Sn[,k]+BSn[2])
  for (j in seq(2,NbPts-2,1))
    {SnQQ[,k]<-SnQQ[,k]+(as.numeric(Sn[,k]>=PtsQQTn[j,2] & Sn[,k]<PtsQQTn[j+1,2]))*
    (ASn[j+1]*Sn[,k]+BSn[j+1])
  }
  SnQQ[,k]<-SnQQ[,k]+(as.numeric(Sn[,k]>=PtsQQTn[NbPts-1,2]))*(ASn[NbPts]*Sn[,k]+BSn[NbPts])
}

# fusion avec correction QQ et ré-écriture fichier ordonné

nomfichier<-paste(Chemin,"STn8015QQ.txt",sep="")
write.table(cbind(Sn,SnQQ), file=nomfichier, sep=",",
  col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA")

# On s'occupe ensuite de Tx
#####

...

# Intercomparaison de statistiques climatologiques classiques
#####

...
# Partie 1/3 sur les données observées

...

# Partie 2/3 sur les données simulées en milieu de période (1998)

...

# Partie 3/3 sur les données simulées sur l'ensemble de la période

...

```

En sortie, on obtient tout d'abord des impressions de contrôle comme la liste des points pivots qui se présente sous cette forme (exemple pour des températures minimales Tn) :

	IND	SMOY	TN
[1,]	2	-14.8843902	-19.1601543
[2,]	8	-12.8738632	-14.6714107
[3,]	10	-12.4802788	-14.4601543
[4,]	18	-11.5912403	-12.9601543
[5,]	28	-10.9038395	-11.7601543
[6,]	46	-10.0856275	-11.0601543
[7,]	74	-9.2492730	-9.6714107
[8,]	120	-8.3272876	-8.6601543
[9,]	194	-7.3593947	-7.6601543
[10,]	314	-6.3075838	-6.2601543
[11,]	508	-5.1581342	-5.1714107
[12,]	822	-3.8586207	-3.8714107
[13,]	1330	-2.3652614	-2.2714107
[14,]	2152	-0.5932038	-0.4601543
[15,]	3482	1.6668834	1.6398457
[16,]	5634	4.9012910	5.0285893
[17,]	7516	7.7861471	7.7916098
[18,]	9668	11.0566568	10.9285893
[19,]	10998	13.0749385	13.0398457
[20,]	11820	14.5025909	14.5398457
[21,]	12328	15.6216428	15.6398457
[22,]	12642	16.5649901	16.5398457
[23,]	12836	17.3821846	17.3285893
[24,]	12956	18.1175550	18.0398457
[25,]	13030	18.7768744	18.7285893
[26,]	13076	19.3976198	19.3285893
[27,]	13104	19.9708494	19.9285893
[28,]	13122	20.5382531	20.3285893
[29,]	13132	21.0146699	20.8285893
[30,]	13140	21.6271876	21.3285893
[31,]	13142	21.8792245	21.7285893
[32,]	13148	23.2125172	22.3285893

En colonnes se trouvent le numéro d'ordre du point, puis son classement (IND parmi les valeurs rangées par ordre croissant), puis la valeur obtenue sur la moyenne de l'ensemble des simulations (SMOY) et enfin la valeur obtenue sur l'observation (TN).

Un ensemble de simulations parfait, ne nécessitant pas de correction post-simulation, afficherait des valeurs de SMOY égales à celles de TN.

D'une manière générale, les écarts entre SMOY et la variable observée sont les plus faibles pour TX, ne devenant significatifs que pour les très basses valeurs. Ils sont un peu plus importants pour TN, avec une plage de conformité plus restreinte. Ceci s'accroît encore avec l'ETP, ce qui est normal car cette opération (de correction quantile-quantile) permet aussi de gérer le cas des valeurs négatives qui n'existent pas dans les observations.

On obtient ensuite la liste des 30 écarts les plus importants (en valeur absolue) restant entre SMOY corrigé et la variable observée :

	Point	SnMoyQQ	TNxxxxx	DeltaAbs
[1,]	13149	22.659005	23.939846	1.2808412
[2,]	4	-16.817164	-15.960154	0.8570100
[3,]	5	-16.178046	-15.560154	0.6178912
[4,]	19	-12.822274	-12.460154	0.3621197
[5,]	7	-15.113730	-15.460154	0.3464241
[6,]	56	-10.508013	-10.171411	0.3366022
[7,]	51	-10.774398	-10.460154	0.3142433
[8,]	55	-10.559590	-10.260154	0.2994357
[9,]	57	-10.445472	-10.171411	0.2740616
[10,]	52	-10.715531	-10.460154	0.2553771
[11,]	54	-10.609418	-10.360154	0.2492636
[12,]	23	-12.316880	-12.071411	0.2454693
[13,]	14	-13.644026	-13.408390	0.2356361
[14,]	48	-10.936139	-10.708390	0.2277489
[15,]	20	-12.681045	-12.460154	0.2208909
[16,]	58	-10.391308	-10.171411	0.2198974
[17,]	49	-10.881934	-10.671411	0.2105230
[18,]	63	-10.171064	-9.971411	0.1996529
[19,]	53	-10.658464	-10.460154	0.1983099
[20,]	21	-12.566728	-12.371411	0.1953177
[21,]	24	-12.203093	-12.008390	0.1947029
[22,]	61	-10.258088	-10.071411	0.1866778
[23,]	147	-8.255485	-8.071411	0.1840746
[24,]	59	-10.349223	-10.171411	0.1778128
[25,]	69	-9.886031	-9.708390	0.1776410
[26,]	68	-9.935853	-9.760154	0.1756991
[27,]	22	-12.434724	-12.260154	0.1745702
[28,]	50	-10.832726	-10.660154	0.1725712
[29,]	110	-8.840935	-8.671411	0.1695247
[30,]	70	-9.840858	-9.671411	0.1694473

On peut remarquer qu'il n'y a aucun point pivot dans cette liste puisque la correction est, par construction, parfaite en ces points.

Et il y a enfin un affichage de statistiques climatiques (Série observée 1980-2015, simulées *100 en 1998, simulées*100 en moyenne sur 1980-2015) dont voici un extrait :

```
# Partie 1/3 sur les données observées
[1] "Txnnnnn    Jours où Tx>=30°C"
[1] "NbObs      : 36"
[1] "Minimum    : 2"
[1] "Maximum    : 50"
[1] "Moyenne    : 14.5556"
[1] "Ec.-Type   : 9.1603"

# Partie 2/3 sur les données simulées en milieu de période (1998)
[1] "Txnnnnn    Jours où Tx>=30°C"
[1] "NbSimul    : 100"
[1] "Minimum    : 0"
[1] "Maximum    : 32"
[1] "Moyenne    : 14.93"
[1] "Ec.-Type   : 6.597"

# Partie 3/3 sur les données simulées sur l'ensemble de la période
[1] "Txnnnnn    Jours où Tx>=30°C"
[1] "NbSimul    : 100"
[1] "Minimum    : 12.0556"
[1] "Maximum    : 17.1667"
[1] "Moyenne    : 14.6072"
[1] "Ec.-Type   : 0.9961"
```

Il est important de vérifier que :

- Les moyennes observées et simulées sont très proches
- L'écart-type observé dans les simulations au centre de la période est légèrement inférieur à celui des observations (car une partie de la variance des valeurs dépend de l'avancée du temps dans le cadre du Changement Climatique)
- L'écart-type de la moyenne des simulations sur 1980-2015 multiplié par la racine carrée du nombre de simulations vaut bien l'écart-type de la moyenne observée.

Une incongruité manifeste dans ces sorties signifie qu'il y a eu une erreur de manipulation. Ceci devrait rester exceptionnel dans le cadre des procédures automatisées. C'était assez fréquent dans les procédures manuelles, à cause du grand nombre de reports d'écriture.

En plus de ces sorties imprimées, vous récupérez aussi 38 fichiers pour les températures et 56 avec les ETP dont seuls les fichiers de configurations des points pivots sont indispensables pour la seconde exécution des projections et sont à laisser sur place. Il s'agit des fichiers « PtsQQTn.txt », « PtsQQTx.txt » et « PtsQQETP.txt ».

(Utilisation du fichier de commande SimTxxxxxQQ10M.R ou SimETPTxxxxxQQ10M.R)

Le suffixe « QQ10M » signifie avec correction Quantile-Quantile et en 10 000 exemplaires. Il ne comporte plus la recherche des points pivot pour correction QQ mais une simple exécution finale de cette correction.

Je suggère un premier passage « pour essai » avec 10 simulations pour vérifier qu'il n'y a pas de plantage avant de lancer le paquet des 10 000 simulations.

```
# Simulation de Boen sur Lignon (Tnx 42019)
# sur période 01/01/1980 - 31/12/2050, soit 25933 jours consécutifs.
# Ecriture fichiers sur période 01/01/1980 - 31/12/2050, pour statistiques.
# Ecriture fichier public sur 01/01/1980 - 31/12/2050 arrondi à 2 décimales
# pour limite de taille sous systèmes 32 bits.

# Récupération des coefficients d'organisation par fichiers. Résidus finaux gaussiens.
# Initialisation par fichiers des caractéristiques (moy,sd) des résidus et des premiers pas
# de la reconstruction de court-terme.
# Correction QQ post-simulation sur liste de valeurs à indices fixes, calcul parallélisé.

# Traitement des valeurs manquantes, facultatif si déjà inclus à l'installation de R
options(na.action = "na.exclude")

# Initialisations

Chemin<-"D:/Rfiles/SimulationTnx/R1/"
NumStat<-"42019"

SauvTxt<-FALSE (notez qu'on ne sauvegarde plus que les fichiers au format binaire, sauf fichiers finaux)
DebSauv<-19800101
DebSauvPub<-19800101
NbPas<-25933
NbSimu<-10000 (premier passage avec 10, par sécurité)

...
# On traite le cas d'un résidu final gaussien

# Création des tableaux de résidus avec variance uniforme
#####

...
# Création des résidus avec variance dé-uniformisée
#####

...
# Création des résidus auto-corrélés, ajout du schéma d'autocorrélation
#####

...
# Création des températures simulées, ajout de l'organisation de long terme
#####

...
# Application de la correction Quantile-Quantile finale
#####
```

```

...
# Correction pour Tn
# Calcul des paramètres des segments de correction
# Importation des coordonnées des points

nomfichier<-paste(Chemin,"PtsQQTn.txt",sep="")
PtsQQTn <- as.matrix(
  read.table(file=nomfichier,header=TRUE, sep="," , na.strings="NA", dec=".", strip.white=TRUE))

NbPts<-nrow(PtsQQTn)
ASn<-matrix(0,ncol=1,nrow=NbPts)
BSn<-matrix(0,ncol=1,nrow=NbPts)

for (i in seq(2,NbPts,1))
  {ASn[i]<-(PtsQQTn[i,3]-PtsQQTn[i-1,3])/(PtsQQTn[i,2]-PtsQQTn[i-1,2])
  BSn[i]<-PtsQQTn[i-1,3]-ASn[i]*PtsQQTn[i-1,2]
  }

# Application de la formule par segments (calcul parallélisé)

for (k in seq(1,NbSimu,1))
  {SnQQ[,k]<-SnQQ[,k]+(as.numeric(Sn[,k]<PtsQQTn[2,2]))*(ASn[2]*Sn[,k]+BSn[2])
  for (j in seq(2,NbPts-2,1))
    {SnQQ[,k]<-SnQQ[,k]+(as.numeric(Sn[,k]>PtsQQTn[j,2] & Sn[,k]<PtsQQTn[j+1,2]))*
    (ASn[j+1]*Sn[,k]+BSn[j+1])
    }
  SnQQ[,k]<-SnQQ[,k]+(as.numeric(Sn[,k]>PtsQQTn[NbPts-1,2]))*(ASn[NbPts]*Sn[,k]+BSn[NbPts])
  }

# Correction pour Tx
...

# Ecriture fichiers finaux arrondis à 2 décimales et sur période limitée

SnQQ<-round(SnQQ,2)
SxQQ<-round(SxQQ,2)

nomfichier<-paste(Chemin,"Sn",NumStat,"Pub.txt",sep="")
write.table(subset(cbind(DateJ,SnQQ),(DateJ>=DebSauvPub)),
  file=nomfichier, sep="," ,
  col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA")

nomfichier<-paste(Chemin,"Sx",NumStat,"Pub.txt",sep="")
write.table(subset(cbind(DateJ,SxQQ),(DateJ>=DebSauvPub)),
  file=nomfichier, sep="," ,
  col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA")

```

En sortie, on obtient essentiellement des fichiers de projections. Parmi les 12 fichiers pour les températures (18 en ajoutant l'ETP), seuls les 4 derniers (6 avec l'ETP) sont de stockage indispensable. Il s'agit des projections finales, avec correction Quantile-Quantile, dont les versions texte se présentent toujours de la même manière, soit une première colonne de datage au format AAAAMMJJ et 10 000 colonnes suivantes contenant les 10 000 séries projetées (entre le 01/01/1980 et le 31/12/2050 dans notre exemple).

Attention à la taille de ces fichiers, de l'ordre de 1,3Go pour chaque fichier texte et de 1,7Go pour chaque fichier binaire au format « R ». L'intérêt de ces fichiers binaires est un accès environ 3 fois plus rapide lors de la lecture sous procédure « R », ce qui peut être sensible pour des masses de données aussi importantes.

5.4. Précipitations

Les procédures correspondantes comprennent un passage supplémentaire par rapport à l'ETP ou aux températures.

De plus, la première étape nécessite une opération supplémentaire. En effet, en plus du transfert vers l'espace de travail « R » des 6 fichiers de configuration obtenus lors de l'analyse d'organisation, il faut également y transmettre les projections de températures de la station thermique de référence (celle utilisée pour l'analyse d'organisation), et ceci au format numérique .Rda.

Attention à bien utiliser (lorsqu'elles sont disponibles) les projections de températures obtenues conjointement avec celles de l'ETP, faute de quoi les variables RR et ETP ne seront pas en

concordance immédiate, ce qui empêcherait la plupart des utilisations de variables combinées, comme le Bilan Hydrique Potentiel (RR-ETP).

(Utilisation du fichier de commande SimLRRxxxx.R)

On choisit un exemple de fichier où la station thermique est déportée par rapport à la station pluviométrique, le cas local étant plus simple. On déroule d'abord les 3 étapes du remontage des organisations : les résidus de variance uniforme (gaussiens), les résidus avec variance organisée (ajout de l'organisation de la variance), et enfin les valeurs simulées complètes (ajout de l'organisation de court terme à l'aide des projections de températures déjà réalisées).

```
# Simulation de Chateauneuf la Forêt (LRR 87040) (inscription manuelle à volonté)
# sur période 01/01/1980 - 31/12/2050, soit 25933 jours consécutifs.
# Ecriture fichiers sur période 01/01/1980 - 31/12/2050, pour statistiques.
# Ecriture fichier public sur 01/01/1980 - 31/12/2050 arrondi à 2 décimales
# pour limite de taille sous systèmes 32 bits.

# Récupération des coefficients d'organisation par fichiers. Résidus finaux gaussiens.
# Initialisation par fichiers des caractéristiques (moy,sd) des résidus et des premiers pas
# de la reconstruction de court-terme. Version avec Tnx déportées.
# Correction QQ post-simulation sur liste de valeurs à indices fixes, calcul parallélisé.

# Initialisations

Chemin<-"D:/Rfiles/SimulationRR/R9/" (mettre manuellement le répertoire de travail, les sous-répertoires
R1,R2,...Ri sont utiles pour des exécutions multiples et parallèles de plusieurs stations différentes)
NumStatRR<-"87040" (mettre manuellement le numéro INSEE de la station pluviométrique)
NumStatT<-"87034" (mettre manuellement le numéro INSEE de la station thermique)

SauvTxt<-TRUE (sortie optionnelle de fichiers intermédiaires, étape par étape de construction)
DebSauv<-19800101 (mettre le début des observations au format AAAAMJJ)
DebSauvPub<-19800101 (idem)
NbPas<-25933 (nombre de jours entre le début des observations et la fin des projections)
NbSimul<-100 (c'est un minimum, mais on pourrait avoir le même nombre que pour la deuxième exécution)

...

# On traite le cas d'un résidu final gaussien

# Création des tableaux de résidus avec variance uniforme
#####

# Pour LRR, on écrit le cas d'un résidu gaussien avec expression directe
# récupération des moyennes et écart-types précalculés
nomfichier<-paste(Chemin,"MoySDLRRSaETu",NumStatRR,".txt",sep="")
MoySDLRR<-
read.table(file=nomfichier,header=TRUE,sep=" ",dec=".")
MoyenneLRR<-MoySDLRR$MOY;ETypeLRR<-MoySDLRR$SD

SLRSaETu <- matrix(rnorm(NbPas*NbSimul, mean=MoyenneLRR, sd=ETypeLRR),ncol=NbSimul)
colnames(SLRSaETu) <- paste("SLR", 1:NbSimul, sep="")

# Ecriture format texte (optionnelle)
if(SauvTxt)
  {nomfichier<-paste(Chemin,"SLRSaETu.txt",sep="")
  write.table(subset(cbind(DateJ,SLRSaETu),(DateJ>=DebSauv)),
  file=nomfichier, sep=" ",
  col.names=TRUE, row.names=FALSE, quote=TRUE, na="NA")}

# Ecriture format R Data, compact
SLRSaETuRda<-cbind(DateJ,SLRSaETu)
nomfichier<-paste(Chemin,"SLRSaETu.Rda",sep="")
save(SLRSaETuRda,file=nomfichier)
rm(SLRSaETuRda)

# Création des résidus avec variance dé-uniformisée
#####

# Récupération des coefficients d'organisation de la variance

...

# Remplissage des tableaux
for(j in seq(1,NbSimul,1))
  {for(i in seq(1,NbPas,1))
    {SLRSauLVO[i,j]<-(SLRSaETu[i,j])/LRRC*(LRRC+LRR Tend*(i+15)/36524
    +LRR Amp1*sin(2*acos(-1)*(i+15)/365.25+LRR Phas1)
    +LRR Amp2*sin(2*acos(-1)*(i+15)/182.625+LRR Phas2))
    }
  }

...

```

```

# Création des résidus auto-corrélés, ajout du schéma d'autocorrélation
#####

# Récupération des coefficients d'organisation de court terme
...

# Remplissage des tableaux de Tn et Tx par récupération des simulations antérieures
(spécifique aux précipitations !)

nomfichier<-paste (Chemin, "Sn", NumStatT, "QQ.Rda", sep="")
load(file=nomfichier)
nomfichier<-paste (Chemin, "Sx", NumStatT, "QQ.Rda", sep="")
load(file=nomfichier)

for(j in seq(1, NbSimul, 1))
  {Sn[,j]<-SnQQRda[,j+1]
  Sx[,j]<-SxQQRda[,j+1]}

# Remplissage des premiers éléments (non autocorrélés) des tableaux
# récupération des moyennes et écart-types précalculés
nomfichier<-paste (Chemin, "MoySDLRR", NumStatRR, ".txt", sep="")
MoySDLRR<-
read.table (file=nomfichier, header=TRUE, sep=" ", dec=".")
MoyenneLRR<-MoySDLRR$MOY; ETypeLRR<-MoySDLRR$SD

for(j in seq(1, NbSimul, 1))
  {for(i in seq(1, 7, 1))
    {SLR[i,j]<-rnorm(1, mean=MoyenneLRR, sd=ETypeLRR)
    }
  }

# Remplissage des éléments suivants
for(i in seq(8, NbPas, 1))
  {SLR[i,j]<- (LRRRC+LRRRTend*i/36524+LRRRAmp1*sin(2*acos(-1)*i/365.25+LRRPhas1)+LRRRAmp2*sin(2*acos(-1)*i/
182.625+LRRPhas2)

+Sx[i,j]*(LRRB0C+LRRB0Tend*i/36524+LRRB0Amp1*sin(2*acos(-1)*i/365.25+LRRB0Phas1)+LRRB0Amp2*sin(2*acos(-1)*i/
182.625+LRRB0Phas2))
+Sx[i-1,j]*(LRRB1C+LRRB1Amp1*sin(2*acos(-1)*i/365.25+LRRB1Phas1)+LRRB1Amp2*sin(2*acos(-1)*i/
182.625+LRRB1Phas2))
+Sx[i-2,j]*(LRRB2C+LRRB2Amp1*sin(2*acos(-1)*i/365.25+LRRB2Phas1))
+Sx[i-3,j]*(LRRB3C+LRRB3Amp1*sin(2*acos(-1)*i/365.25+LRRB3Phas1)+LRRB3Amp2*sin(2*acos(-1)*i/182.625+LRRB3Phas2))
+Sx[i-4,j]*(LRRB4C+LRRB4Amp1*sin(2*acos(-1)*i/365.25+LRRB4Phas1)+LRRB4Amp2*sin(2*acos(-1)*i/182.625+LRRB4Phas2))

+SLR[i-1,j]*(LRRRC1C+LRRRC1Tend*i/36524+LRRRC1Amp1*sin(2*acos(-1)*i/365.25+LRRRC1Phas1)+LRRRC1Amp2*sin(2*acos(-1)*i/
182.625+LRRRC1Phas2))
+SLR[i-2,j]*(LRRRC2C+LRRRC2Amp1*sin(2*acos(-1)*i/365.25+LRRRC2Phas1))
+SLR[i-3,j]*(LRRRC3C)
+SLR[i-6,j]*(LRRRC6C+LRRRC6Amp1*sin(2*acos(-1)*i/365.25+LRRRC6Phas1))
+SLRSauLVO[i,j])
  }
}
...

# Attention, ici l'organisation de long terme des LRR est obtenue implicitement par celle des Tn et Tx
#####
...

```

On déroule ensuite les opérations permettant d'effectuer la correction quantile-quantile post-simulation, mais en se focalisant sur les aspects spécifiques des précipitations.

```

#####
# Ensemble de procédures de préparation à la projection des précipitations avec correction QQ
#####
(Ce qui suit est donc spécifique au premier passage de la procédure de projection climatique)

# Initialisations : date de début, date de fin des obs
# liste des numéros de points ordonnés pivots pour correction QQ

DebObs<-19800101; FinObs<-20151231 (attention, il faut écrire ces dates manuellement)
AnDeb<-DebObs %/% 10000; AnFin<-FinObs %/% 10000
IndicFibo<-c(2, 6, 12, 20, 30, 45, 60, 80, 110, 150, 210, 300, 400, 500, 650, 800, 1000, 1250, 1500, 1750, 2000,
2400, 2800, 3200, 3600, 4000, 4500, 5000, 5500, 6000, 6500, 6650, 7150, 7650, 8150, 8650, 9150,
9550, 9950, 10350, 10750, 11150, 11400, 11650, 11900, 12150, 12350, 12500, 12650, 12750,
12850, 12940, 13000, 13040, 13070, 13090, 13105, 13120, 13130, 13138, 13144, 13148)
NbPts<-NROW(IndicFibo)
(On a conservé le même nom de variable « Fibo » que pour les températures et l'ETP mais on n'a plus respecté
une suite de Fibonacci qui donne un espacement trop lâche pour les valeurs centrales de la fonction de
répartition. Il faut une discrétisation plus fine.)

# On commence par la fusion des simulations avec l'observation, sur la période observée
#####
...

# On poursuit par le tri des valeurs colonne par colonne
#####

```



```

...
# Création des fichiers de points pivots pour correction QQ
#####
...
# Applications de la correction QQ sur échantillon de projections (pour contrôle)
#####
# PARTIE TRAITEMENT DES FICHIERS ORDONNES #
#####
...
# Impression de contrôle des trente résidus de correction QQ les plus élevés
...
# PARTIE TRAITEMENT DES FICHIERS NON-ORDONNES #
#####
...
# Intercomparaison de statistiques climatologiques classiques
#####
...

```

En sortie, on obtient tout d'abord des impressions de contrôle comme la liste des points pivots qui se présente sous cette forme :

	IND	SMOY	LRR
[1,]	2	-3.34470083	-1.99900731
[2,]	6	-3.06807621	-1.99657709
[3,]	12	-2.91023864	-1.99385140
...			
[25,]	3600	-0.99382242	-1.00245947
[26,]	4000	-0.92740216	-0.95870554
[27,]	4500	-0.84903470	-0.91345092
[28,]	5000	-0.77380355	-0.87289846
[29,]	5500	-0.70126399	-0.83348601
[30,]	6000	-0.62892990	-0.79820385
[31,]	6500	-0.55735443	-0.76614974
[32,]	6650	-0.53594798	-0.75750693
[33,]	7150	-0.46414639	-0.72847995
[34,]	7650	-0.39223319	-0.70059369
[35,]	8150	-0.31853680	-0.63559764
[36,]	8650	-0.24191893	-0.44722981
[37,]	9150	-0.16312037	-0.27746750
[38,]	9550	-0.09559019	-0.14792843
[39,]	9950	-0.02463590	-0.00825177
[40,]	10350	0.05123183	0.13325497
[41,]	10750	0.13468088	0.27133249
[42,]	11150	0.22697660	0.40593851
[43,]	11400	0.29179103	0.48499578
[44,]	11650	0.36414710	0.58186133
[45,]	11900	0.44534283	0.67317686
...			
[57,]	13105	1.54728432	1.54852204
[58,]	13120	1.65607642	1.58435915
[59,]	13130	1.75443094	1.63068379
[60,]	13138	1.87698452	1.68136469
[61,]	13144	2.05101324	1.72553539
[62,]	13148	2.32528045	1.76711693

On peut déjà remarquer que, par rapport aux autres paramètres, les différences entre les quantiles de la moyenne des simulations et ceux de la série observée des précipitations sont plus importantes et surtout plus instables entre quantiles successifs. C'est la raison pour laquelle le nombre de points pivots pour la correction quantile-quantile est passé de 32 à 62. Ce nombre, et l'espacement entre quantiles, résultent de nombreux essais.

On obtient ensuite la liste des 30 écarts les plus importants (en valeur absolue) restant entre SMOY corrigé et la variable observée :

```

print(TabDelta[1:30,])
      Point  SLRMoyQQ  LRR63113  DeltaAbs
[1,]  8071 -0.6459103 -0.6779385  0.03202815
[2,]  8072 -0.6457841 -0.6777939  0.03200978
[3,]  8073 -0.6456496 -0.6776049  0.03195534
[4,]  8070 -0.6460332 -0.6779490  0.03191586
[5,]  8069 -0.6461513 -0.6780169  0.03186563
[6,]  8068 -0.6462633 -0.6780283  0.03176502
[7,]  8074 -0.6455338 -0.6772784  0.03174455
[8,]  8067 -0.6463894 -0.6780765  0.03168707
[9,]  8066 -0.6465587 -0.6780885  0.03152988
[10,] 8076 -0.6452883 -0.6766974  0.03140910
...
[26,] 8053 -0.6482866 -0.6786306  0.03034407
[27,] 8047 -0.6490655 -0.6793312  0.03026573
[28,] 8046 -0.6491852 -0.6794454  0.03026020
[29,] 8043 -0.6495758 -0.6798181  0.03024224
[30,] 8079 -0.6448916 -0.6751258  0.03023423

```

Il est particulièrement significatif que la totalité de ces 30 écarts les plus importants soient concentrés autour d'un quantile quasi central. C'est une illustration de l'instabilité des écarts, dont les conséquences ont pu être limitées (ici à 0,03 soit 3 % d'erreur sur les précipitations) grâce à l'augmentation du nombre de points pivots.

Il y a un équilibre à obtenir entre l'augmentation du nombre de points pivots (au plus on en met, au plus la qualité statistique de détermination se dégrade) et les erreurs résiduelles. On peut probablement encore un peu progresser, en augmentant localement la densité des points pivots, là où les erreurs systématiques les plus importantes se sont régulièrement produites. Pour l'ensemble des stations, ces écarts systématiques (jusque 15 % d'erreur dans les cas les plus sévères) se sont produits entre les observations ordonnées n°7000 et n°11500. Si on ne l'a pas fait ici, c'est que ces erreurs relatives maximales se produisent sur des valeurs de précipitation quotidienne de quelques dixièmes de mm, avec une influence négligeable sur les cumuls.

Et il y a enfin un affichage de statistiques climatiques (Série observée 1980-2015, simulées *100 en 1998, simulées*100 en moyenne sur 1980-2015) dont voici un extrait :

```

# Partie 1/3 sur les données observées
[1] "RRxxxxx   Jours où RR>=1mm"
[1] "NbObs     : 36"
[1] "Minimum   : 62"
[1] "Maximum   : 111"
[1] "Moyenne   : 84.6667"
[1] "Ec.-Type  : 11.8683"

# Partie 2/3 sur les données simulées en milieu de période (1998)
[1] "RRxxxxx   Jours où RR>=1mm"
[1] "NbSimul   : 100"
[1] "Minimum   : 66"
[1] "Maximum   : 107"
[1] "Moyenne   : 84.18"
[1] "Ec.-Type  : 8.2455"

# Partie 3/3 sur les données simulées sur l'ensemble de la période
[1] "RRxxxxx   Jours où RR>=1mm"
[1] "NbSimul   : 100"
[1] "Minimum   : 80.8889"
[1] "Maximum   : 89.3333"
[1] "Moyenne   : 84.5728"
[1] "Ec.-Type  : 1.5397"

```

mêmes remarques que pour les températures et l'ETP.

En plus de ces sorties imprimées, vous récupérez aussi 13 fichiers dont seul le fichier de configuration des points pivots est indispensable pour la deuxième et la troisième exécution des projections et est à laisser sur place. Il s'agit du fichier « PtsQLRR.txt ».

(Première utilisation du fichier de commande SimLRRxxxxxQQ10M.R)

On fera un premier passage « pour essai » avec 10 simulations pour vérifier qu'il n'y a pas de plantage avant de lancer le paquet des 10 000 simulations.

```
# Simulation de Chateauneuf la Forêt (LRR 87040)
# sur période 01/01/1980 - 31/12/2050, soit 25933 jours consécutifs.
# Ecriture fichiers sur période 01/01/1980 - 31/12/2050, pour statistiques.
# Ecriture fichier public sur 01/01/1980 - 31/12/2050 arrondi à 2 décimales
# pour limite de taille sous systèmes 32 bits.

# Récupération des coefficients d'organisation par fichiers. Résidus finaux gaussiens.
# Initialisation par fichiers des caractéristiques (moy,sd) des résidus et des premiers pas
# de la reconstruction de court-terme. Version avec Tnx déportées.
# Correction QQ post-simulation sur liste de valeurs à indices fixes, calcul parallélisé.

# Initialisations

Chemin<-"D:/Rfiles/SimulationRR/R9/"
NumStatRR<-"87040"
NumStatT<-"87034"

SauvTxt<-FALSE (notez qu'on ne sauvegarde plus que les fichiers au format binaire, sauf fichiers finaux)
DebSauv<-19800101
DebSauvPub<-19800101
NbPas<-25933
NbSimul<-10000 (premier passage avec 10, par sécurité)

...
# On traite le cas d'un résidu final gaussien

# Création des tableaux de résidus avec variance uniforme
#####

...
# Création des résidus avec variance dé-uniformisée
#####

...
# Création des résidus auto-corrélés, ajout du schéma d'autocorrélation
#####

...
# Attention, ici l'organisation de long terme des LRR est obtenue implicitement par celle des Tn et Tx
#####

# Application de la correction Quantile-Quantile finale
#####

...
```

La portion de code qui est présentée juste après cette portion de texte mérite toute votre attention. Elle est opérationnelle à la seconde utilisation de la procédure en cours *SimLRRxxxxxQQ10M.R*. Elle correspond à une opération supplémentaire, par rapport aux températures et à l'ETP, de recalibration des projections de précipitations.

Elle s'est imposée à cause de la manière dont l'organisation de long terme des précipitations a dû être considérée, par l'intermédiaire de celles des températures et via un schéma de court-terme. Le problème du schéma de court-terme utilisé seul, c'est que les valeurs de l'instant t sont calculées à partir des valeurs des instants t-1 à t-7 et qu'on doit recommencer cette opération pour le pas de temps suivant. Une éventuelle dérive infime au départ se propage donc et grossit jusqu'à devenir évidente, avec des évolutions incohérentes pour environ 80 % du réseau des stations, à l'échéance 2050.

Perdre autant de stations n'était évidemment pas acceptable. Il a donc fallu se résoudre à imposer une contrainte externe (décidée humainement) pour recalibrer les projections de précipitations. Ceci a donné lieu à un nombre très important d'essais dans une optique de moindre action, c'est-à-dire que la contrainte imposée soit la plus « faible » possible, tout en supprimant les évolutions absurdes.

Cette contrainte minimale c'est *in fine* le maintien de la tendance sur le cumul annuel entre la période observée 1980-2015 et la période projetée 2000-2050, ceci s'entend sur la variable non-transformée en logarithme. Pour limiter l'incertitude statistique, ce n'est pas directement la tendance 1980-2015 de la station elle-même qui est utilisée mais une interpolation géographique de ce

paramètre, qu'on peut lire par exemple sur une cartographie basique, mais qu'on pourrait aussi calculer par une pondération des stations limitrophes en fonction inverse de leur distance. La « tendance » sur les projections est calculée par différence entre l'espérance (moyenne des 10 000 simulations) de l'échéance 2050 et celle de l'échéance 2000. La correction est nulle en 2000, car le procédé du générateur stochastique de temps garantit la bonne ré-analyse des paramètres climatiques au centre de la période observée (2000≈1998), et elle s'accroît ensuite linéairement jusque 2050, pour redresser la pente.

Dans la première utilisation de cette procédure (soit ici même!), cette partie de code est rendue sans effet par le traitement de l'absence de fichier de configuration.

Différentes pistes pourraient être explorées pour supprimer cette nécessité de recalibration ad-hoc, parmi lesquelles ne rien faire et attendre ! En effet, d'après la littérature, le temps d'émergence d'un signal significatif sur le paramètre précipitations (c'est-à-dire le délai à partir duquel l'évolution de long terme de ce paramètre est statistiquement significative) serait de 50 à 75ans, à partir du moment où l'évolution significative a débuté. Si on admet qu'elle a débuté en même temps que celle des températures, c'est-à-dire vers 1975, alors on pourrait y être presque (2025), mais dans d'autres endroits il faudrait attendre une génération (2050). Il n'est pas impossible qu'on y soit déjà si on s'intéresse au Sud-Ouest de la France où l'évolution est déjà nette à la baisse.

Lorsqu'on se trouvera dans ce cas, il suffira d'adopter, pour les précipitations, le même type de schéma d'organisation que celui des températures et de l'ETP. On pourrait même effectuer l'analyse d'organisation (et les projections) conjointement entre ces 3 types de paramètre.

```
# Opération supplémentaire de recalibration externe relative, sur tendance linéaire cartographique lissée en 50
ans
# Cette recalibration s'applique sur la variable RR=10^LRR-0.11.
# Elle est unitaire au 30/06/2000 (i=7487) et atteint la valeur cible au 30/06/2050 (i=25748).

# récupération des valeurs de recalibration, avec valeurs de non-calibration au besoin
nomfichier<-paste(Chemin,"RecalRR.txt",sep="")
if (file.exists(nomfichier))
{RecalRR<-
read.table(file=nomfichier,header=TRUE,sep=" ",dec=".")
RRAn2000Cible<-RecalRR$RR00Cible;DeltaCible<-RecalRR$DCible;RRAn2050Val<-RecalRR$RR50Val
}else
{RRAn2000Cible<-1000
DeltaCible<- (0)
RRAn2050Val<-1000
}
RRAn2050Cible<-RRAn2000Cible+DeltaCible

for(j in seq(1,NbSimul,1))
{For(i in seq(1,NbPas,1))
{RR<-10^(SLRQQ[i,j])-0.11
Alpha<-(RRAn2050Cible/RRAn2050Val-1)/18261;Beta<-(1-7487*Alpha)
RapCor<-Alpha*i+Beta
RRModif<-max(RR*RapCor,(-0.109999))
SLRQQ[i,j]<-log10(RRModif+0.11)
}
}
...

```

La procédure se termine par une écriture des fichiers finaux. Faire attention aux 3 possibilités de format pour les précipitations : le format binaire et le format texte pour la variable de base LRR et aussi un format texte supplémentaire en variable ordinaire RR, pour un usage direct des précipitations non-transformées.

```
# Ecriture des fichiers de sortie finaux, variable LRR format binaire
...
# Ecriture fichier final en variable RR, arrondi à 2 décimales et sur période limitée
...
# Ecriture fichier final en variable LRR arrondi à 3 décimales et sur période limitée
...

```

Pour cette première utilisation, le seul fichier de sortie qui nous intéresse est le fichier final texte en variable RR. Il permet en effet le calcul du cumul moyen des précipitations en échéance 2000 et en échéance 2050. A ces deux valeurs, il faut ajouter la détermination de l'évolution cible, pondérée géographiquement, et inclure les trois dans un fichier de configuration de la recalibration appelé « RecalRR.txt » qui se présente sous cette forme :

```
RR00Cible DCible RR50Val  
771.8 -5 986.1
```

Les séparations entre données sont des espaces. On ne change surtout rien à la première ligne et, pour la seconde, on écrit dans l'ordre le cumul moyen annuel des projections de précipitations en échéance 2000, l'évolution cible et le cumul moyen annuel des projections de précipitations en échéance 2050.

On pourrait être tenté de réduire le nombre de simulations de ce premier passage (à 1000 par exemple) à cause des temps d'exécution qui se chiffrent en heures mais la variabilité statistique est importante. Le procédé nécessite effectivement les 10 000 simulations pour stabiliser les valeurs des cumuls moyens de précipitations.

(Seconde utilisation du fichier de commande SimLRRxxxxxQQ10M.R)

Comme il n'y a aucune modification de code, on peut lancer directement le paquet des 10 000 simulations. Le listing n'est pas rappelé puisqu'il est identique.

En sortie, on obtient essentiellement des fichiers de projections. Parmi les 6 fichiers, seuls les 3 derniers sont de stockage indispensable. Il s'agit des projections finales, après correction quantile-quantile, de format identique à celui des températures et de l'ETP.

Les projections étant terminées, il reste à voir deux exemples d'utilisation de ces projections, qui ont été effectivement développés lors du projet AP3C, à savoir des calculs d'Indicateurs Agro-Climatiques (IAC) et des calculs de Bilan Hydrique Réel (BHR).

6. Les Indicateurs Agro-Climatiques (IAC)

Un indicateur agroclimatique est une variable composée d'éléments climatiques qui ont été déterminés ou/et qui ont été articulés pour correspondre à des besoins ou des problématiques agricoles.

On verra d'abord la catégorisation des IAC, tout au moins ceux calculés dans le cadre de AP3C, puis on détaillera le fonctionnement de quelques procédures de calcul.

6.1. Catégorisation

Commençons par rappeler quelques éléments de contexte. L'obtention des projections climatiques du projet AP3C a été un processus prototype qui s'est étagé de l'automne 2016, pour le démarrage des températures, au printemps 2018, pour la fin des précipitations.

Cette durée « extrême » de 18 mois, intervenant de plus une année après le démarrage du projet, posait problème quant à sa continuité dans les autres composantes, en particulier pour maintenir l'implication des agronomes des chambres d'agriculture départementales participantes. Les IAC constituant les premiers résultats spécifiquement agronomiques, une première salve de 12 indicateurs purement thermiques a été calculée dès le début de l'année 2017. Un second paquet de 18 indicateurs, pouvant contenir en plus de la pluviométrie et de l'ETP, a été calculé durant l'été et le début d'automne 2018.

Ceci nous amène à la première nomenclature, effectivement utilisée dans les résumés agronomiques départementaux, qui est un numéro d'ordre chronologique IAC_{nn} où nn varie de 01 à 12 pour les indicateurs purement thermiques et de 13 à 30 pour les autres. Il faut bien reconnaître que cet inventaire n'a pas de pertinence fonctionnelle.

On a donc mis au point une seconde nomenclature, par type de végétal plus une catégorie générale indifférenciée, en deux lettres suivies d'un numéro d'ordre :

Prnn : Prairie (18 IAC) Cénn : Céréales (6 IAC) Msn : Maïs (6 IAC)
 Dénn : Dérobées (2 IAC) Génn : Générale (2 IAC) Vinn : Vigne (1 IAC)

On remarquera que cette seconde catégorisation permet d'intégrer les résultats de 5 IAC supplémentaires, qui n'ont pas été inclus dans les publications soit par ce qu'il s'agissait de modes de calcul alternatifs soit de procédures auxiliaires.

On présente ci-après 2 tableaux récapitulatifs de ces 35 IAC. Voici, tout d'abord, un tableau spécifique aux 18 IAC relatifs à la prairie :

Code fonctionnel	Code chrono.	Intitulé	Formule de calcul	Paramètres complémentaires
Pr00	X	Prairie : Cumul de température en janvier	Somme de températures	Base : 0°C – Max : 18°C, initialisée au 01/01
Pr01	IAC01	Prairie : Date de démarrage de végétation	Date à laquelle on atteint une somme de températures de 200°Cj	Base : 0°C – Max : 18°C, initialisée au 01/01
Pr02	IAC02	Prairie : Date de mise à l'herbe	Date à laquelle on atteint une somme de températures de 250°Cj	Base : 0°C – Max : 18°C, initialisée au 01/02
Pr03	IAC03	Prairie : Date de fauche précoce	Date à laquelle on atteint une somme de températures de 750°Cj	Base : 0°C – Max : 18°C, initialisée au 01/02
Pr04	IAC04	Prairie : Date de fauche intermédiaire	Date à laquelle on atteint une somme de températures de 1000°Cj	Base : 0°C – Max : 18°C, initialisée au 01/02
Pr05	IAC05	Prairie : Date de fauche tardive	Date à laquelle on atteint une somme de températures de 1200°Cj	Base : 0°C – Max : 18°C, initialisée au 01/02

Pr06	IAC13	Prairie : Satisfaction hydrique après mise à l'herbe	Calcul du ratio RR/ETP. Période calée sur sommes de températures.	Seuils 250°Cj et 500°CJ (initialisés au 01/02)
Pr07	IAC14	Prairie : Satisfaction hydrique avant l'ensilage	Calcul du ratio RR/ETP. Période calée sur sommes de températures.	Seuils 500°Cj et 750°CJ (initialisés au 01/02)
Pr08	IAC15	Prairie : Satisfaction hydrique de l'ensilage à la récolte en foin	Calcul du ratio RR/ETP. Période calée sur sommes de températures.	Seuils 750°Cj et 1000°CJ (initialisés au 01/02)
Pr09	IAC16	Prairie : Séquences favorables à la récolte en ensilage	Nombre de jours tels que les 2 jours précédents soient sans pluie (RR<1mm chaque jour) et que le cumul des 5 jours encore avant soit inférieur à 20mm.	Période de 15 jours centrée sur date cumul 750°Cj , soit date ± 7 jours
Pr10	IAC17	Prairie : Séquences favorables à la récolte en foin sec	Nombre de jours tels que les 4 jours précédents soient sans pluie (RR<1mm chaque jour) et que le cumul des 5 jours encore avant soit inférieur à 20mm.	Période de 21 jours centrée sur date cumul 1100°Cj , soit date ± 10 jours
Pr11	IAC18	Prairie : Satisfaction hydrique en automne	Calcul du ratio RR/ETP. Période fixe.	Du 15/09 au 30/10
Pr12	IAC25	Prairie : Satisfaction hydrique estivale après récolte en foin	Calcul du ratio RR/ETP. Période avec début fixe et fin calée sur somme de températures.	Seuil de début de période 1200°Cj, fin 15/09
Pr13	IAC26	Prairie : Réussite hydrique des semis d'automne	Cumul de précipitations. Période avec début fixe et fin calée en anticipation statistique de gel fort.	Du 15/08 à la date antérieure de 20 jours au décile n°2 de la date de première gelée à -5°C
Pr14	IAC27	Prairie : Satisfaction hydrique hivernale	Calcul du ratio RR/ETP. Période fixe.	Du 30/10 au 31/12
Pr15	IAC28	Prairie : Séquences favorables à la récolte en enrubannage	Nombre de jours tels que les 3 jours précédents soient sans pluie (RR<1mm chaque jour) et que le cumul des 5 jours encore avant soit inférieur à 20mm.	Période de 15 jours centrée sur date cumul 750°Cj , soit date ± 7 jours
Pr16	IAC29	Prairie : Périodes favorables à la mise en place des semis de prairies de printemps	Nombre de jours tels que les 5 jours précédents soient sans pluie (RR<1mm chaque jour)	Période de 21 jours démarrant sur date cumul 250°Cj (initialisée au 01/02)
Pr17	X	Prairie : Précipitations entre mise à l'herbe et fauche précoce	Cumul de précipitations. Période calée sur sommes de températures.	Seuils 250°Cj et 750°CJ (initialisés au 01/02)

Tableau 1 : Liste des IAC relatifs à la prairie

Et voici le tableau relatif aux 17 autres indicateurs :

Code fonctionnel	Code chrono.	Intitulé	Formule de calcul	Paramètres complémentaires
Cé01	IAC06	Céréales : Risque de gel au stade épi 1cm	Nombre de jours avec température minimale inférieure à un seuil. Période fixe.	Seuil : -4°C. Du 20/02 au 10/04
Cé02	IAC07	Céréales : Intensité de l'échaudage	Nombre de jours avec température maximale supérieure à un seuil. Période fixe.	Seuil : 25°C. Du 15/05 au 20/07

Cé03	IAC23	Céréales : Précipitations lors du remplissage du grain. En altitude	Cumul de précipitations. Période fixe.	Du 10/06 au 30/06
Cé04	IAC24	Céréales : Précipitations lors du remplissage du grain. En plaine	Cumul de précipitations. Période fixe.	Du 20/05 au 10/06
Cé05	X	Céréales : Réussite de l'apport azoté au stade épi 1cm. En altitude	Nombre de jours avec cumul quotidien de précipitations supérieur à un seuil. Période fixe.	Seuil : 15mm. Du 01/04 au 15/04
Cé06	X	Céréales : Réussite de l'apport azoté au stade épi 1cm. En plaine	Nombre de jours avec cumul quotidien de précipitations supérieur à un seuil. Période fixe.	Seuil : 15mm. Du 15/03 au 30/03
Ms01	IAC08	Maïs : Echaudage thermique	Nombre de jours avec température maximale supérieure à un seuil. Période fixe.	Seuil : 32°C. Du 01/06 au 30/09
Ms02	IAC09	Maïs : Risque de gel en fin de cycle	Première date à laquelle la température minimale est inférieure à un seuil. Période fixe.	Seuil : -2°C. Du 01/09 au 31/10
Ms03	X	Maïs : Risque de gel en fin de cycle (bis)	Nombre de jours avec température minimale inférieure à un seuil. Période fixe.	Seuil : -2°C. Du 01/09 au 31/10
Ms04	IAC19	Maïs : Choix variétaux, gamme d'indices	Somme de températures , période fixe.	Base : 6°C – Max : 30°C, du 10/05 au 15/10
Ms05	IAC20	Maïs : Stress hydrique de la floraison au remplissage du grain	Nombre de décades où le cumul de précipitations est supérieur à un seuil. Période fixe.	Seuil : 20mm. Du 01/07 au 20/08
Ms06	IAC30	Maïs : Stress hydrique de la floraison au remplissage du grain (bis)	Nombre de décades tel que au moins un cumul quotidien est supérieur à un seuil. Période fixe.	Seuil : 10mm. Du 01/07 au 20/08
Dé01	IAC21	Dérobées : Faisabilité thermique des dérobées de printemps	Somme de températures. Période conditionnelle : démarrage sur seuil de somme de températures (herbe) , arrêt sur considérations statistiques de levée avant gel fort.	Base principale 11°C. Base secondaire (herbe) 0°C. Seuil sur sommes herbe 1000°Cj. Arrêt au décile n°2 de la première gelée d'automne à -5°C (soit non gel probable 4/5).
Dé02	IAC22	Dérobées : Faisabilité thermique des dérobées d'été	Somme de températures. Période conditionnelle : démarrage fixe , arrêt sur considérations statistiques de levée avant gel fort.	Base principale 11°C. Du 10/07 au décile n°2 de la date de première gelée d'automne à -5°C (soit non gel probable 4/5).
Gé01	IAC10	Généraliste : Dernière gelée de printemps	Dernière date à laquelle la température minimale est inférieure à un seuil. Période fixe.	Seuil : 0°C. Du 01/01 au 30/06
Gé02	IAC11	Généraliste : Première gelée forte d'automne	Première date à laquelle la température minimale est inférieure à un seuil. Période fixe.	Seuil : -5°C. Du 01/10 au 31/12
Vi01	IAC12	Vigne : Indice héliothermique de Huglin	Résultat de la formule : $\Sigma[(T_m-10)+(T_x-10)]*1,04/2$	Du 01/04 au 30/09

Tableau 2 : Liste des IAC hors prairie

Pour aider au choix des exemples de procédures qui seront présentés juste après, on peut effectuer un autre type de regroupement, sur des considérations fonctionnelles, par type de calcul du plus simple au plus complexe. On pourrait distinguer 6 groupes de procédures :

- Les cumuls ou ratios sur période fixe : Pr00, Pr11, Pr14, Cé03, Cé04, Cé05, Cé06, Ms04 et Vi01.
- Les dénombrements d'événements simples sur période fixe : Cé01, Cé02, Ms01, Ms03, Ms05 et Ms06
- Les premières ou dernières dates d'événements simples sur période fixe : Ms02, Gé01 et Gé02
- Les cumuls ou ratios sur période conditionnelle simple, calée sur des sommes de températures : Pr01, Pr02, Pr03, Pr04, Pr05, Pr06, Pr07, Pr08, Pr12 et Pr17
- Les dénombrements complexes sur période conditionnelle simple : Pr09, Pr10, Pr15 et Pr16
- L'intégration de conditionnalités faisant intervenir des statistiques d'autres indicateurs : Dé01, Dé02 et Pr13

6.2. Exemples de procédures

On choisira un exemple pour chacun des 6 types de procédure. Les listings complets de la totalité des 35 procédures seront bien entendu à disposition dans l'annexe « Fichier Projections IAC.ods ».

6.2.1. Cumul ou ratio sur période fixe

On passe en revue la procédure relative à l'indicateur Ms04 : « choix variétaux, gamme d'indices » .

L'architecture générale des procédures est toujours la même, pour plus de répliquabilité. Après le titre, on spécifie la séparation générale entre une première partie de procédure relative aux projections et la seconde relative aux observations. Ces deux parties ont été conçues pour être les plus semblables possibles, de sorte qu'une modification de l'une soit très facilement reportable sur l'autre.

```
# Indicateur agro-climatique Ms04, mais choix de variété
#####

# Création de statistiques climatologiques sur cumuls de température
#####

# PARTIE SIMULATIONS #
#####
...

```

Le premier bloc actif est très important puisqu'il s'agit des initialisations, qu'il faut réaliser à la main. Les initialisations, ou paramétrisations, ont été rendues les plus nombreuses possibles pour pouvoir facilement changer de date, de base de sommes de températures, de seuils divers...

```
# Initialisations, nombre de simulations étudiées, échéances, numéros station,
# date de début et de fin des cumuls, bornes min et max du cumul

NomIndic<-"Ms04"
Chemin<="/home/debian/Documents/Rfiles/ProjectionIAC/"
NbSimul<-10000
Echeances<-1980:2050
NbEcheances<-NROW(Echeances)
NumStat<-c("03060","03155","03180","03248","03285","12145","12214","12254","12300","15014","15053",
"15114","15122","15187","15226","19031","19036","19146","19237","192754","23001","23008",
"23013","23021","23025","23030","23031","23045","23067","23075","23079","23093","230961",
"230967","230968","23106","23118","23142","23176",
"42005","42019","42170","43062","43096","43111","43130","43234","46005","46083","46127",
"46160","46197","48039","48051","48070","48104","48116","48140","63003","63345","63354",
```

```

      "87009", "870324", "87044", "870856", "87089", "87187")
NbStat<-NROW(NumStat)
DateDeb<-510;DateFin<-1015
BaseMin<-6;BaseMax<-30
...

```

On passe ensuite à l'écriture des 2 tableaux de réception des résultats. Le premier tableau comprend le calcul de l'indicateur année par année et simulation par simulation, le second intègre des statistiques basiques de ces calculs.

```

# Création des tableaux de réception des résultats
# Somme de Tm, à la date de fin de cumul
SomTm<-matrix(0,nrow=NROW(Echeances),ncol=NbSimul)
colnames(SomTm)<-paste("Sim",1:NbSimul,sep="")
rownames(SomTm)<-Echeances

# Statistiques sur somme de Tm
StatSomTm<-matrix(0,nrow=NROW(Echeances),ncol=13)
colnames(StatSomTm)<-c("P05",paste("Déc",1:9,sep=""),"P95","Moy","SD")
rownames(StatSomTm)<-Echeances
...

```

On démarre le calcul de l'indicateur par la récupération des projections climatiques dans des tableaux locaux, ici de températures minimales et maximales.

```

# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{

# Lecture des fichiers d'entrée.

# On s'occupe d'abord des températures minimales Tn
nomfichier<-paste(Chemin,"Sn",NumStat[IndicStat],"Pub.txt",sep="")
Dataset <- as.matrix(
  read.table(file=nomfichier,
    header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE))

# On ajoute l'année et le mois-jour au fichier de NbSimul+1 colonnes
Date<-Dataset[,1]
An<-Date %/% 10000
MoisJour<-(Date-An*10000)
SimTn<-cbind(Dataset,An,MoisJour)

# On s'occupe ensuite des températures maximales Tx
nomfichier<-paste(Chemin,"Sx",NumStat[IndicStat],"Pub.txt",sep="")
Dataset <- as.matrix(
  read.table(file=nomfichier,
    header=TRUE, sep=" ", na.strings="NA", dec=".", strip.white=TRUE))

SimTx<-cbind(Dataset,An,MoisJour)

# On élimine ce qui est inutile
rm(Dataset)

...

```

Parfois, il faut transformer les variables de base, comme ici lorsqu'on les borne par un mini et un maxi.

```

# On calcule les températures bornées entre BaseMin et BaseMax (6 et 30°C classiquement)

SimTnB6<-matrix(0,nrow=nrow(SimTn),ncol=NbSimul+3)
SimTxB6<-matrix(0,nrow=nrow(SimTx),ncol=NbSimul+3)

# Bornage avec les fonctions parallélisables
VectSix<-rep(BaseMin,NbSimul)
VectTrente<-rep(BaseMax,NbSimul)
for (i in seq(1,nrow(SimTn),1))
{
  SimTnB6[i,2:(NbSimul+1)]<-pmax(VectSix,SimTn[i,2:(NbSimul+1)])
  SimTnB6[i,2:(NbSimul+1)]<-pmin(VectTrente,SimTnB6[i,2:(NbSimul+1)])
  SimTxB6[i,2:(NbSimul+1)]<-pmax(VectSix,SimTx[i,2:(NbSimul+1)])
  SimTxB6[i,2:(NbSimul+1)]<-pmin(VectTrente,SimTxB6[i,2:(NbSimul+1)])
}

rm(SimTn);rm(SimTx)

# On termine par la moyenne et on ajoute les 3 colonnes de repérage temporel
SimTmB6<-(SimTnB6+SimTxB6)/2-BaseMin
SimTmB6[,1]<-Date
SimTmB6[,NbSimul+2]<-An
SimTmB6[,NbSimul+3]<-MoisJour

# Libération de mémoire

```

```
rm(SimTnB6);rm(SimTxB6)
...
```

Voici le (petit) cœur du calcul, ici un simple cumul d'une température moyenne bornée. A la fin, l'indicateur est extrait du tableau de calcul et stocké dans le tableau de sortie.

```
for (i in seq(1,NbEcheances,1))
# Calcul de cumul total année par année pour T
{TmB6An<-subset(SimTmB6,An==Echeances[i])
# On met le cumul à zéro
CumTmB6An<-TmB6An*0
CumTmB6An[,1]<-TmB6An[,1]
CumTmB6An[,NbSimul+2]<-TmB6An[,NbSimul+2]
CumTmB6An[,NbSimul+3]<-TmB6An[,NbSimul+3]

for (j in seq(which(TmB6An[,NbSimul+3]==DateDeb),nrow(TmB6An),1))
{CumTmB6An[j,2:(NbSimul+1)]<-CumTmB6An[j-1,2:(NbSimul+1)]+TmB6An[j,2:(NbSimul+1)]}

# Calcul du cumul à la date de fin
for (k in seq(2,NbSimul+1,1))
{NumJourFin<-which(CumTmB6An[,NbSimul+3]==DateFin)
SomTm[i,k-1]<-CumTmB6An[NumJourFin,k]
}
}
```

Il reste à remplir le tableau de statistiques à l'aide du tableau de sortie, simulation par simulation. Veuillez noter que le calcul des quantiles, moyenne et écart-type ne se fait pas directement sur la totalité des 10 000 valeurs mais sur 100×100 (car $\sqrt{10\,000}=100$). C'est assez superflu pour un indicateur continu mais indispensable pour les indicateurs « quantiques », qui ne peuvent prendre que des valeurs entières, surtout quand ces valeurs peuvent être proches de zéro. On passerait ainsi directement de 1 à 0 ou à 2... difficile de cerner des évolutions non-radicales dans ce cas. D'où l'astuce d'utiliser une forme de transitivité des quantiles, calculés sur des sous-échantillons (de 100 simulations) puis agrégés en moyenne de 100 paquets. Pour ceux qui souhaiteraient aller plus loin, la transitivité n'est exacte qu'avec la moyenne. Il y a des déformations (une faible contraction vers les valeurs centrales) avec les quantiles, surtout les plus extrêmes, qu'il conviendrait d'étudier de manière académique et éventuellement de corriger.

```
# Calculs de statistiques par agrégation de sous-échantillons
# Quantiles par paquets de (NbSimul^0,5)
Quant<-0
for (j in seq(1,sqrt(NbSimul),1))
{Quant<-Quant+quantile(SomTm[i,((j-1)*sqrt(NbSimul)+1):(j*sqrt(NbSimul))],probs=0.05)/sqrt(NbSimul)}
StatSomTm[i,1]<-Quant

for (k in seq(2,10,1))
{Quant<-0
for (j in seq(1,sqrt(NbSimul),1))
{Quant<-Quant+quantile(SomTm[i,((j-1)*sqrt(NbSimul)+1):(j*sqrt(NbSimul))],probs=(k-1)*0.1)/sqrt(NbSimul)}
StatSomTm[i,k]<-Quant
}

Quant<-0
for (j in seq(1,sqrt(NbSimul),1))
{Quant<-Quant+quantile(SomTm[i,((j-1)*sqrt(NbSimul)+1):(j*sqrt(NbSimul))],probs=0.95)/sqrt(NbSimul)}
StatSomTm[i,11]<-Quant

Moy<-0
for (j in seq(1,sqrt(NbSimul),1))
{Moy<-Moy+mean(SomTm[i,((j-1)*sqrt(NbSimul)+1):(j*sqrt(NbSimul))])/sqrt(NbSimul)}
StatSomTm[i,12]<-Moy

EType<-0
for (j in seq(1,sqrt(NbSimul),1))
{EType<-EType+sd(SomTm[i,((j-1)*sqrt(NbSimul)+1):(j*sqrt(NbSimul))])/sqrt(NbSimul)}
StatSomTm[i,13]<-EType
}
```

On procède à l'écriture des résultats et on passe à la seconde partie relative aux observations. Cette partie constitue une sorte de marque de fabrique de AP3C qui s'est engagé, et a été en partie fondé, sur la représentation simultanée (sur un même document graphique) de ce qui est relatif aux observations et de ce qui est relatif aux projections. Il y a quelques initialisations complémentaires, qui ne sont donc pas au début de la procédure mais en début de seconde partie (attention!).

```

# Ecriture des résultats complets puis des statistiques
nomfichier<-paste(Chemin,"IAC_",NomIndic,"_",NumStat[IndicStat],".txt",sep="")
write.table(round(SomTm,0),file=nomfichier,sep="," ,
  col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")

nomfichier<-paste(Chemin,"Stats_",NomIndic,"_",NumStat[IndicStat],".txt",sep="")
write.table(round(StatSomTm,1),file=nomfichier,sep="," ,
  col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
}

```

```

# PARTIE OBSERVATIONS #
#####

```

```

# Initialisations spécifiques

NomFicObsT<-paste(Chemin,"TnxMap3cHV.txt",sep="")

```

```

Echeances<-1980:2015
NbEcheances<-NROW(Echeances)
TnNom<-paste("TN",NumStat,"Hmv",sep="")
TxNom<-paste("TX",NumStat,"Hmv",sep="")

```

```

SomTmObs<-matrix(0,nrow=NROW(Echeances),ncol=1)
colnames(SomTmObs)<-paste("Obs","_Hmv",sep="")
rownames(SomTmObs)←Echeances
...

```

Notez la particularité d'importation du fichier d'observations commun à toutes les stations, et aussi aux températures minimales et maximales ! Ceci a contraint à simuler une importation interne supplémentaire depuis le tableau général vers des tableaux qui sont semblables à ceux des projections mais avec un nombre de « simulation » égal à un.

```

# Importation des fichiers obs globaux
#####

TableauT <- as.matrix(
  read.table(file=NomFicObsT,
    header=TRUE, sep="," , na.strings="NA", dec=".", strip.white=TRUE))

```

```

# On déroule la procédure station après station
#####

```

```

for (IndicStat in seq(1,NbStat,1))
{
  # Création des "fichiers" d'entrée.
  # On s'occupe d'abord des températures minimales Tn
  Dataset <- TableauT[,c("DATE",TnNom[IndicStat])]

  # On ajoute l'année et le mois-jour au fichier de NbSimul+1 colonnes
  Date<-Dataset[,1]
  An<-Date %/% 10000
  MoisJour<-(Date-An*10000)
  ObsTn<-cbind(Dataset,An,MoisJour)

  # On s'occupe ensuite des températures maximales Tx
  Dataset <- TableauT[,c("DATE",TxNom[IndicStat])]
  ObsTx<-cbind(Dataset,An,MoisJour)

  # On élimine ce qui est inutile
  rm(Dataset)
  ...
}

```

Maintenant, on peut dérouler les mêmes instructions que pour les projections.

```

# On calcule les températures bornées entre BaseMin et BaseMax (6 et 30°C classiquement)

```

```

ObsTnB6<-matrix(0,nrow=nrow(ObsTn),ncol=4)
ObsTxB6<-matrix(0,nrow=nrow(ObsTx),ncol=4)

# Bornage avec les fonctions parallélisables (un peu superflues ici...)
VectSix<-rep(BaseMin,1)
VectTrente<-rep(BaseMax,1)
for (i in seq(1,nrow(ObsTn),1))
{ObsTnB6[i,2:2]<-pmax(VectSix,ObsTn[i,2:2])
  ObsTnB6[i,2:2]<-pmin(VectTrente,ObsTnB6[i,2:2])
  ObsTxB6[i,2:2]<-pmax(VectSix,ObsTx[i,2:2])
  ObsTxB6[i,2:2]<-pmin(VectTrente,ObsTxB6[i,2:2])}

rm(ObsTn);rm(ObsTx)

```

```

# On termine par la moyenne et on ajoute les 3 colonnes de repérage temporel
ObsTmB6<-(ObsTnB6+ObsTxB6)/2-BaseMin
ObsTmB6[,1]<-Date
ObsTmB6[,3]<-An
ObsTmB6[,4]<-MoisJour

```

```

# Libération de mémoire
rm(ObsTnB6);rm(ObsTxB6)

for (i in seq(1,NbEcheances,1))

# Calcul de cumul total année par année pour T
{TmB6An<-subset(ObsTmB6,An==Echeances[i])
# On met le cumul à zéro
CumTmB6An<-TmB6An*0
CumTmB6An[,1]<-TmB6An[,1]
CumTmB6An[,3]<-TmB6An[,3]
CumTmB6An[,4]<-TmB6An[,4]

for (j in seq(which(TmB6An[,4]==DateDeb),nrow(TmB6An),1))
{CumTmB6An[j,2:2]<-CumTmB6An[j-1,2:2]+TmB6An[j,2:2]}

# Calcul du cumul à la date de fin
for (k in seq(2,2,1))
{NumJourFin<-which(CumTmB6An[,4]==DateFin)
SomTmObs[i,k-1]<-CumTmB6An[NumJourFin,k]
}
}
...

```

Bien sûr, il n'y a pas de calculs de statistiques puisqu'il n'y a qu'une série observée pour le calcul de l'indicateur. On stocke simplement le calcul direct.

```

# Ecriture des résultats
nomfichier<-paste(Chemin,"Obs_",NomIndic,"_",NumStat[IndicStat],".txt",sep="")
write.table (round(SomTmObs,0),file=nomfichier,sep="," ,
col.names=TRUE, row.names=TRUE, quote=TRUE, na="NA")
}

```

Pour les autres exemples qui vont suivre, on ne reprendra pas la totalité du listing de la procédure, mais on se focalisera sur les particularités supplémentaires ou alternatives.

6.2.2. Evénements simples sur période fixe

On s'intéresse à la procédure relative à l'indicateur Cé02 : « intensité de l'échaudage » .

```

# Indicateur agro-climatique Cé02, céréales intensité de l'échaudage thermique
#####

# Création de statistiques climatologiques sur nombre de jours
#####

# PARTIE SIMULATIONS #
#####
...

```

Concernant les initialisations, notez simplement la possibilité (non-utilisée ici) de pouvoir décaler de manière prescrite et progressive les dates de calcul.

```

# Initialisations, nombre de simulations étudiées, échéances, numéros station, seuil de franchissement
# dates de début et de fin de sensibilité, décrétement annuel de dates
...

DateDeb<-515
DateFin<-720
DecAn<-0.0
SeuilTx<-(25)

...

# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{
...

```

Le cœur du calcul est assez simple : détermination dynamique des dates de début et fin de période (qui maintient les valeurs fixes ici) puis vérification de condition ($T_x > 25^\circ\text{C}$) et

accumulation dans une variable locale avant inscription dans le tableau général de réception des résultats.

```
# Calcul de nombre de jours de dépassement de seuil, année par année

for (i in seq(1,NbEcheances,1))
  {TxAn<-subset(SimTx,An==Echeances[i])
# Numéro des jours de démarrage et de fin, avec décalage éventuel
  NumJourD<-which(TxAn[,NbSimul+3]==DateDeb)
  NumJourD<-NumJourD-round((i-1)*DecAn,0)
  NumJourF<-which(TxAn[,NbSimul+3]==DateFin)
  NumJourF<-NumJourF-round((i-1)*DecAn,0)

# Calcul du nb de jours de dépassement de seuil, simulation par simulation
  for (k in seq(2,NbSimul+1,1))
    {NbJSeuil<-0
      for (j in seq(NumJourF,NumJourD,-1))
        {if (TxAn[j,k]>=SeuilTx)
          {NbJSeuil<-NbJSeuil+1}
        }
      NbJSupTx[i,k-1]<-NbJSeuil
    }

# Calculs de statistiques par agrégation de sous-échantillons

...

# Ecriture des résultats complets puis des statistiques
nomfichier<-paste(Chemin,"IAC_",NomIndic,"_",NumStat[IndicStat],".txt",sep="")
...

# PARTIE OBSERVATIONS #
#####

...

```

Rien de particulier concernant la partie traitant des observations.

6.2.3. Datation d'un événement, sur période fixe

On s'intéresse à la procédure relative à l'indicateur Gé01 : « dernière gelée de printemps » .

```
# Indicateur agro-climatique Gé01, généraliste gel tardif de printemps
#####

# Création de statistiques climatologiques sur date de dernier dépassement d'un seuil de température
#####

# PARTIE SIMULATIONS #
#####

# Initialisations, nombre de simulations étudiées, échéances, numéros station, seuil de franchissement
# date maximale de fin de printemps

...
DateDeb<-SGelPrin<-(0)
DateFinPrin<-630
...

# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{
# Lecture des fichiers d'entrée.
...

# Calcul des dates de dépassement de seuil, année par année
...
}

```

Notez qu'on initialise la date du dernier gel au 1^{er} janvier, qui est donc la date par défaut s'il n'y a aucun gel du 01/01 (codé « 101 ») au 30/06 (codé « 630 »). Comme l'opération est simple (une date plus tardive écrase la précédente), on inscrit le résultat directement dans le tableau de réception, sans passer par une variable locale. La « date » en question est en fait le numéro du jour de l'année, qui est une variable discrète ordinaire sur laquelle on peut effectuer des statistiques.

```
for (i in seq(1,NbEcheances,1))
  {TnAn<-subset(SimTn,An==Echeances[i])
# Recherche de dates de franchissement simulation par simulation

```

```

for (k in seq(2,NbSimul+1,1))
  {DateGelPrin[i,k-1]<-which(TnAn[,NbSimul+3]==101)
  for (NumJour in seq(which(TnAn[,NbSimul+3]==101),which(TnAn[,NbSimul+3]==DateFinPrin),1))
    {if (TnAn[NumJour,k]<=SGelPrin)
      {DateGelPrin[i,k-1]<-NumJour}
    }
  }

# Calculs de statistiques par aggrégation de sous-échantillons
...
# Ecriture des résultats complets puis des statistiques
...

# PARTIE OBSERVATIONS #
#####
...

```

Rien de particulier concernant la partie traitant des observations.

6.2.4. Cumul ou ratio sur période conditionnelle simple (sommés de températures)

On s'intéresse à la procédure relative à l'indicateur Pr03 : « date de fauche précoce » .

```

# Indicateur agro-climatique Pr03, prairie fauche précoce
#####

# Création de statistiques climatologiques sur date de franchissement de seuil
#####

# PARTIE SIMULATIONS #
#####

# Initialisations, nombre de simulations étudiées, échéances, numéros station, seuil de franchissement
# et date de début de cumul

...
SeuilSomT<-750
DateDebCum<-201
...
# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{

# Lecture des fichiers d'entrée.
...

```

Attention, comme cette procédure faisait partie des premières développées, le bornage sur les températures n'est pas paramétrisé (initialisé). Les bornes 0°C et 18°C ont été codées « en dur ».

```

# On calcule les températures bornées entre 0 et 18°C

SimTnB0<-matrix(0,nrow=nrow(SimTn),ncol=NbSimul+3)
SimTxB0<-matrix(0,nrow=nrow(SimTx),ncol=NbSimul+3)

# Bornage avec les fonctions parallélisables
VectZero<-rep(0,NbSimul)
VectDixHuit<-rep(18,NbSimul)
for (i in seq(1,nrow(SimTn),1))
  {SimTnB0[i,2:(NbSimul+1)]<-pmax(VectZero,SimTn[i,2:(NbSimul+1)])
  SimTnB0[i,2:(NbSimul+1)]<-pmin(VectDixHuit,SimTnB0[i,2:(NbSimul+1)])
  SimTxB0[i,2:(NbSimul+1)]<-pmax(VectZero,SimTx[i,2:(NbSimul+1)])
  SimTxB0[i,2:(NbSimul+1)]<-pmin(VectDixHuit,SimTxB0[i,2:(NbSimul+1)])}
...

```

Le calcul se fait en deux étapes, d'abord un cumul sur toutes les dates à partir de la date de début puis une recherche de la première date où ce cumul dépasse le seuil de somme de températures prescrit.

```

# Calcul de cumul année par année

for (i in seq(1,NbEcheances,1))
  {TmB0An<-subset(SimTmB0,An==Echeances[i])
# On met le cumul à zéro
CumTmB0An<-TmB0An*0
CumTmB0An[,1]<-TmB0An[,1]
CumTmB0An[,NbSimul+2]<-TmB0An[,NbSimul+2]
CumTmB0An[,NbSimul+3]<-TmB0An[,NbSimul+3]

for (j in seq(which(TmB0An[,NbSimul+3]==DateDebCum),nrow(TmB0An),1))
  {CumTmB0An[j,2:(NbSimul+1)]<-CumTmB0An[j-1,2:(NbSimul+1)]+TmB0An[j,2:(NbSimul+1)]}

```

```

# Recherche de date de franchissement simulation par simulation
for (k in seq(2,NbSimul+1,1))
  {DateF<-0;NumJourF<-0
  for (j in seq (nrow(TmB0An),1,-1))
    {if (CumTmB0An[j,k]>=SeuilSomT)
      {DateF<-CumTmB0An[j,NbSimul+3]}
    }
  NumJourF<-which(CumTmB0An[,NbSimul+3]==DateF)
  NumJFSeuil[i,k-1]<-NumJourF
}

# Calculs de statistiques par agrégation de sous-échantillons
...

# PARTIE OBSERVATIONS #
#####
...

```

6.2.5. Conditions complexes sur période conditionnelle

On s'intéresse à la procédure relative à l'indicateur Pr10 : « séquences favorables à la récolte en foin sec » .

```

# Indicateur agro-climatique Pr10, Séquences sèches favorables à la récolte du foin
#####

# Création de statistiques climatologiques sur nombres de jours favorables
#####

# PARTIE SIMULATIONS #
#####

# Initialisations, nombre de simulations étudiées, échéances, numéros station, seuil de somme
# de températures, date de début de cumul de température, demi-diamètre de la fenêtre de travaux,
# taille de la période antérieure et seuil de sécheresse, taille de période pré-antérieure et
# seuil de cumul de précipitations

...
SeuilSomT<-1100;DateDebCum<-201
DemiLargPer0<-10
LargPer1<-4;SeuilPer1<-1.0
LargPer2<-5;SeuilCumPer2<-20.0

...
# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{

# Lecture des fichiers d'entrée.
...

```

On récupère les températures pour le calcul auxiliaire de sommes de températures et les précipitations pour l'étude des conditions de faisabilité.

```

# On s'occupe d'abord des températures minimales Tn
...
# On s'occupe ensuite des températures maximales Tx
...
# On s'occupe ensuite des précipitations RR
...

for (i in seq(1,NbEcheances,1))
...

```

Les premiers calculs sont des sommes de températures suivis de la recherche de la date correspondant au dépassement de seuil (ça on connaît déjà, voir procédure précédente).

```

# Calcul de cumul total année par année pour T
{TmB0An<-subset(SimTmB0,An==Echeances[i])
# On met le cumul à zéro
CumTmB0An<-TmB0An*0
CumTmB0An[,1]<-TmB0An[,1]
CumTmB0An[,NbSimul+2]<-TmB0An[,NbSimul+2]
CumTmB0An[,NbSimul+3]<-TmB0An[,NbSimul+3]

for (j in seq(which(TmB0An[,NbSimul+3]==DateDebCum),nrow(TmB0An),1))
  {CumTmB0An[j,2:(NbSimul+1)]<-CumTmB0An[j-1,2:(NbSimul+1)]+TmB0An[j,2:(NbSimul+1)]}

# Calcul du cumul total année après année pour RR
RRAn<-subset(SimRR,An==Echeances[i])

```



```

# On met le cumul à zéro
CumRRAn<-RRAn*0
CumRRAn[,1]<-RRAn[,1]
CumRRAn[,NbSimul+2]<-RRAn[,NbSimul+2]
CumRRAn[,NbSimul+3]<-RRAn[,NbSimul+3]

for (j in seq(which(RRAn[,NbSimul+3]==DateDebCum),nrow(RRAn),1))
  {CumRRAn[j,2:(NbSimul+1)]<-CumRRAn[j-1,2:(NbSimul+1)]+RRAn[j,2:(NbSimul+1)]}

# Recherche de la date de franchissement de la somme de T simulation par simulation
for (k in seq(2,NbSimul+1,1))
  {DateBase<-0;NumJourBase<-0
  for (j in seq(nrow(TmB0An),1,-1))
    {if (CumTmB0An[j,k]>=SeuilSomT)
      {DateBase<-CumTmB0An[j,NbSimul+3]}
    }
  NumJourBase<-which(CumTmB0An[,NbSimul+3]==DateBase)
...

```

On ajoute un niveau dans la complexité en cherchant le nombre de vérifications de critères multiples (critère 1<-> booléen « FavPer1 »...) sur la période déterminée. On passe par une variable locale (JourFav) avant stockage dans le tableau de résultats.

```

# Vérification des critères de jour favorable
JourFav<-0
for (j in seq(NumJourBase-DemiLargPer0,NumJourBase+DemiLargPer0,1))
  {FavPer1<-TRUE
  for (m in seq(1,LargPer1,1))
    {FavPer1<-FavPer1&(RRAn[j-m,k]<SeuilPer1)}
  FavPer2<-(CumRRAn[j-LargPer1-1,k]-CumRRAn[j-LargPer1-LargPer2-1,k])<SeuilCumPer2
  if (FavPer1&FavPer2)
    {JourFav<-JourFav+1}
  }

NbJourFav[i,k-1]<-JourFav
}

```

```

# Calculs de statistiques par agrégation de sous-échantillons
...

```

```

# PARTIE OBSERVATIONS #
#####
...

```

6.2.6. Conditionnalités faisant intervenir les statistiques d'un autre indicateur

On s'intéresse à la procédure relative à l'indicateur Dé01 : « faisabilité thermique des dérobes de printemps » .

```

# Indicateur agro-climatique Dé01, Faisabilité thermique des dérobes de printemps
#####

# Création de statistiques climatologiques sur sommes de température
#####

# PARTIE SIMULATIONS #
#####

# Initialisations, nombre de simulations étudiées, échéances, numéros station,
# bornes min et max du cumul secondaire, seuil et date de départ du cumul secondaire,
# borne min du cumul principal, date de fin du cumul principal

...
BaseMinT0<-0;BaseMaxT0<-18
SeuilSomT0<-1000;DateDebCumT0<-201
BaseMinT1<-11

# Création des tableaux de réception des résultats
...
# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{
...

```

On notera qu'il y a deux bornages distincts, le bornage à 0°C et 18°C sur les sommes de températures auxiliaires et le bornage principal à 11°C. Tous ces bornages sont paramétrisés, donc facilement modifiables.

```

# On calcule les températures bornées entre BaseMin et BaseMax du cumul secondaire
...
# Bornage avec les fonctions parallélisables
VectZero<-rep(BaseMinT0,NbSimul)
VectDixHuit<-rep(BaseMaxT0,NbSimul)
for (i in seq(1,nrow(SimTn),1))
  {SimTnB0[i,2:(NbSimul+1)]<-pmax(VectZero,SimTn[i,2:(NbSimul+1)])
  SimTnB0[i,2:(NbSimul+1)]<-pmin(VectDixHuit,SimTnB0[i,2:(NbSimul+1)])
  SimTxB0[i,2:(NbSimul+1)]<-pmax(VectZero,SimTx[i,2:(NbSimul+1)])
  SimTxB0[i,2:(NbSimul+1)]<-pmin(VectDixHuit,SimTxB0[i,2:(NbSimul+1)])}
...
# On calcule les températures bornées par BaseMin du cumul principal
...
# Bornage avec les fonctions parallélisables
VectOnze<-rep(BaseMinT1,NbSimul)
for (i in seq(1,nrow(SimTn),1))
  {SimTnB11[i,2:(NbSimul+1)]<-pmax(VectOnze,SimTn[i,2:(NbSimul+1)])
  SimTxB11[i,2:(NbSimul+1)]<-pmax(VectOnze,SimTx[i,2:(NbSimul+1)])}
...
for (i in seq(1,NbEcheances,1))
  ...

```

On produit donc 2 cumuls de températures, suivis de la recherche de la première date de dépassement de seuil, pour obtenir la date de démarrage du cumul principal. On inscrit cette date dans une variable locale (« NumJourDebCumT1 »).

```

# Calcul de cumul année par année pour T0 (avec date de début initialisée)
{TmB0An<-subset(SimTmB0,An==Echeances[i])
# On met le cumul à zéro
CumTmB0An<-TmB0An*0
CumTmB0An[,1]<-TmB0An[,1]
CumTmB0An[,NbSimul+2]<-TmB0An[,NbSimul+2]
CumTmB0An[,NbSimul+3]<-TmB0An[,NbSimul+3]

for (j in seq(which(TmB0An[,NbSimul+3]==DateDebCumT0),nrow(TmB0An),1))
  {CumTmB0An[j,2:(NbSimul+1)]<-CumTmB0An[j-1,2:(NbSimul+1)]+TmB0An[j,2:(NbSimul+1)]}

# Calcul de cumul total année par année pour T1
TmB11An<-subset(SimTmB11,An==Echeances[i])
# On met le cumul à zéro
CumTmB11An<-TmB11An*0
CumTmB11An[,1]<-TmB11An[,1]
CumTmB11An[,NbSimul+2]<-TmB11An[,NbSimul+2]
CumTmB11An[,NbSimul+3]<-TmB11An[,NbSimul+3]

CumTmB11An[1,2:(NbSimul+1)]<-TmB11An[1,2:(NbSimul+1)]
for (j in seq(2,nrow(TmB11An),1))
  {CumTmB11An[j,2:(NbSimul+1)]<-CumTmB11An[j-1,2:(NbSimul+1)]+TmB11An[j,2:(NbSimul+1)]}

# Recherche de la date de franchissement de la somme de T0 simulation par simulation
for (k in seq(2,NbSimul+1,1))
  {DateDebCumT1<-0;NumJourDebCumT1<-0
  for (j in seq(nrow(TmB0An),1,-1))
    {if (CumTmB0An[j,k]>=SeuilSomT0)
      {DateDebCumT1<-CumTmB0An[j,NbSimul+3]}
    }
  NumJourDebCumT1<-which(CumTmB0An[,NbSimul+3]==DateDebCumT1)
  ...
}

```

Pour la date de fin du cumul principal, c'est plus complexe. On sait que les statistiques sur les dates de premier gel d'automne à -5°C sont déjà disponibles (procédure G 02), que les ann es indicent ce fichier et que le d cile n 2 est en colonne n 3. Si ces statistiques n'avaient pas  t  pr calcul es, il aurait fallu inclure le code correspondant, ou alors cr er une proc dure compl te pour cet indicateur auxiliaire. Le cumul sur la p riode ainsi d termin e est obtenu par diff rence de cumuls d marrant au 01/01.

```

# Recherche de la date de fin de la somme de T1
# Importation du fichier de statistiques de gel d'automne
nomfichier<-paste(Chemin,"Stats_G 02_",NumStat[IndicStat],".txt",sep="")
StatsGel <- as.matrix(
  read.table(file=nomfichier,
  header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE))

NumAnnee<-which(rownames(StatsGel)==Echeances[i])
NumJourFinCumT1<-round(StatsGel[NumAnnee,3],0)

SomT1[i,k-1]<-(CumTmB11An[NumJourFinCumT1,k]-CumTmB11An[NumJourDebCumT1-1,k])
}

```

```
# Calculs de statistiques par agrégation de sous-échantillons
...
```

```
# PARTIE OBSERVATIONS #
#####
```

```
...
# On déroule la procédure station après station
#####
```

```
for (IndicStat in seq(1,NbStat,1))
{
```

```
...
```

Bien qu'on soit dans la partie sur les données observées, on fait néanmoins appel aux mêmes statistiques que pour les projections, car ces statistiques sont auxiliaires.

```
# Recherche de la date de fin de la somme de T1
# Importation du fichier de statistiques de gel d'automne
nomfichier<-paste(Chemin,"Stats_Gé02_",NumStat[IndicStat],".txt",sep="")
StatsGel <- as.matrix(
read.table(file=nomfichier,
header=TRUE, sep=",", na.strings="NA", dec=".", strip.white=TRUE))

NumAnnee<-which(rownames(StatsGel)==Echeances[i])
NumJourFinCumT1<-round(StatsGel[NumAnnee,3],0)

SomT1Obs[i,k-1]<-(CumTmB11An[NumJourFinCumT1,k]-CumTmB11An[NumJourDebCumT1-1,k])
}
}
...
```

Il n'y a pas vraiment de limite à la complexification de la définition d'un IAC, sauf à vérifier que le concept correspondant est explicitement calculable et que les résultats font sens en agronomie. Les IAC sont essentiellement annuels, mais on pourrait les multiplier sur des sous-périodes pour en faire des valeurs saisonnières, mensuelles ou décadaires.

Ce qui est présenté dans le chapitre suivant va encore plus loin, et ne s'appelle plus IAC, puisqu'on est descendu à une échelle quotidienne combinée à une multiplicité d'indicateurs, dans le cadre d'un modèle de bilan hydrique à deux réservoirs.

7. Le Bilan Hydrique Réel (BHR)

Le terme de « bilan hydrique réel » peut sembler présomptueux dans l'absolu, mais il s'agit d'une expression consacrée qui s'oppose au bilan hydrique potentiel (BHP). Ce BHP est la simple différence entre un cumul de précipitations et une somme d'ETP sur la même période.

Une expression plus précise pourrait être « bilan hydrique à deux réservoirs ». Par rapport à un BHP, ce qu'il y a de plus c'est donc un traitement de l'eau dans le sol, avec des conséquences assez basiques liées à la « profondeur utile », sous entendu utilisable par les racines, et au stress hydrique de la végétation.

Il n'y a pas de traitement de la physiologie du végétal, de la présence ou de l'apport de nutriments, de la présence d'animaux « broutants »... Ce genre de modèle est donc un intermédiaire entre un BHP, purement météorologique, et un véritable modèle agronomique.

On va rappeler son principe général de fonctionnement.

On se définit (à dire d'expert dans AP3C, mais on peut aussi en faire une mesure) une réserve utile en eau du sol (RU) qu'on sépare en deux sous réservoirs. La majeure partie est, en général, la réserve facilement utilisable (RFU) dans laquelle le végétal peut puiser sans stress. Le complément s'appelle la réserve de survie (RS), dans laquelle le végétal peut puiser mais avec un stress progressif, qui devient total lorsque la RS est vide.

Ensuite, on applique un processus séquentiel jour par jour (un modèle « physique » impliquerait un fonctionnement simultané).

On commence par tenter de remplir la réserve utile avec les précipitations. La RFU est d'abord abondée et, si elle déborde, le surplus est mis dans la RS. Si la RS déborde à nouveau, le surplus part en « écoulement ».

On utilise ensuite l'ETP pour calculer une évapotranspiration maximale (ETM) grâce à un coefficient cultural K_c , multiplicatif. Pour l'herbe, par définition, ce coefficient vaut 1, ce qui nous ramène à la valeur de l'ETP. Dans d'autres cas, il peut être notablement différent de 1 et même dépendre sensiblement du stade phénologique ou d'autres considérations (chronologiques, climatiques, génétiques...). On tente ensuite d'enlever l'ETM de la RFU. Si le niveau de la RFU est insuffisant, on tente d'enlever le complément (ETM résiduelle, ou ETMr) de la RS. Mais ici intervient une contrainte supplémentaire liée à la résistance de la RS. L'ETM résiduelle est pondérée par un coefficient multiplicatif, K_s inférieur ou égal à 1, linéairement proportionnel au niveau relatif de la RS par rapport à son niveau maximal. Par exemple, si le niveau de la RS est à 20 % de sa capacité maximale alors K_s vaut 0,2. Ce qu'on enlève in fine de la RS est donc $K_s \cdot ETMr$. La somme de ce qu'on a enlevé de la RFU et de la RS s'appelle l'évapotranspiration réelle (ETR). Le stress hydrique est, par définition, la quantité $1 - (ETR/ETM)$, qu'on exprime généralement en %.

Contrairement à un IAC qui fournit une valeur par an (et par projection climatique), un BHR fournit (basiquement) 5 valeurs par jour : le niveau dans la RFU, le niveau dans la RS, l'écoulement, l'ETR et le stress hydrique. Ça fait environ 1800 fois plus de données et ça ne peut donc pas être traité de la même manière. En particulier, les statistiques quotidiennes ne sont pas très informatives. Il faut plutôt considérer qu'on produit une véritable base de données hydro-pédo-climatiques qui vient s'ajouter à la base strictement climatique des températures, précipitations et ETP.

7.1. Procédure de calcul

Il n'y a qu'une seule procédure de calcul, car il n'y a qu'un seul modèle. Elle est un peu répétitive quant à la gestion des 5 variables hydro-pédo-climatiques, donc on va la dérouler en s'intéressant aux points saillants. La procédure dans son intégralité est à consulter dans l'annexe « Fichier Projections BHR.ods ».

```

# Bilan hydrique réel, herbe toute l'année, ETP observée
#####

# Traitement des valeurs manquantes, facultatif si déjà inclus à l'installation de R
options(na.action = "na.exclude")

# Création d'une base de données d'états de RFU, de RS, d'ETR, de stress hydrique et d'écoulement
#####

# PARTIE SIMULATIONS #
#####
...

```

On notera qu'on se limite fortement en taille de réseau calculé (3 stations ici), à cause de l'importante utilisation de la mémoire vive (au moins 20Go par station). Le coefficient cultural est mis uniformément à 1, car le projet AP3C n'a traité, dans sa phase en cours, que le cas de l'herbe. Pour le sol, on a choisi à dire d'expert un panel de 4 RU (45mm, 75mm, 120mm et 180mm) qu'on a appliqué intégralement à toutes les stations, charge aux utilisateurs de terrain de se rapporter au type de RU qui convient le mieux à leurs parcelles. Notez la répartition en 2/3 pour la RFU et 1/3 pour la RS. Attention, la procédure s'exécute pour un seul type de sol à la fois. Les temps de calcul peuvent devenir imposants, surtout si on n'a pas la capacité de les paralléliser.

```

# Initialisations, nombre de simulations étudiées, numéros station, coefficient cultural,
# niveau maximal de RFU et de RS

NomIndic<-"BH_Herbe"
Chemin<-"D:/Rfiles/ProjectionBHyd/"
NbSimul<-10000
NumStat<-c("03060","03155","03248")

NbStat<-NROW(NumStat)
Kc<-1
RFU0<-30;RU0<-3/2*RFU0;RS0<-(RU0-RFU0)

# On déroule la procédure station après station
#####

for (IndicStat in seq(1,NbStat,1))
{
...

```

On n'utilise en entrée que les données de précipitations et d'ETP.

```

# Lecture des fichiers d'entrée de données
# On s'occupe d'abord des précipitations RR
...
# On s'occupe ensuite des ETP
...

```

Les tableaux pour les statistiques quotidiennes ne sont pas très utiles.

```

# On crée des tableaux de réception des résultats
# état pour chaque simulation et stat pour les statistiques habituelles

```

```

EtatRFU<-Dataset
...

```

On part d'un état initial saturé au 01/01/1980, donc dans certains cas (sols profonds et faibles précipitations en particulier) les résultats seront biaisés la première année. Ça reste intéressant à consulter pour se rendre compte du type d'erreur qu'on commettrait en re-démarrant à saturation hydrique des sols au début de chaque année.

```

# Etat hydrique initial
RFU<-rep(RFU0,NbSimul);RS<-rep(RS0,NbSimul);RU<-rep(RU0,NbSimul)

# Calcul du bilan hydrique réel jour par jour
NbJour<-nrow(SimRR)

for (Jour in seq(1,NbJour,1))
{for (j in seq(1,NbSimul,1))
{
# On traite en premier les précipitations
...

```

Accumulation des précipitations dans la RFU, comparaison à la capacité maximale RFU0, débordement dans la RS et éventuellement évacuation du surplus. Parmi les 5 paramètres quotidiens, l'écoulement peut déjà être traité dans cette partie.

```
RFU[j]<-RFU[j]+SimRR[Jour,j+1]
if (RFU[j]>RFU0)
  {RS[j]<-RS[j]+(RFU[j]-RFU0)
  RFU[j]<-RFU0
  if (RS[j]>RS0)
    {EtatEcoule[Jour,j+1]<-(RS[j]-RS0)
    RS[j]<-RS0
    }else
    {EtatEcoule[Jour,j+1]<-0}
  }else
  {EtatEcoule[Jour,j+1]<-0
  }
```

On traite ensuite l'evapotranspiration

Extraction de l'évaporation. C'est ici que l'ETR et le stress hydrique sont calculés. Les niveaux de RFU et de RS (en fin de journée) sont stockés dans le tableau de sortie, à la suite. Si on envisage de traiter un cas différent de l'herbe permanente, c'est la première ligne qui doit être remplacée, éventuellement par tout un module. En ce cas, il conviendrait sans doute d'inverser l'ordre de description du modèle de BHR, d'abord simulation par simulation et ensuite jour par jour, sauf à transformer Kc en un tableau complet, comme les tableaux de sortie. Pensez à importer les projections et observations de températures si les stades phénologiques qui influencent Kc sont déterminés grâce à des sommes de températures.

```
ETM<-Kc*SimETP[Jour,j+1]
if (RFU[j]>=ETM)
  {RFU[j]<-RFU[j]-ETM
  EtatETR[Jour,j+1]<-ETM
  EtatStressH[Jour,j+1]<-0
  }else
  {ETR1<-RFU[j]
  ETMrest<-ETM-RFU[j]
  RFU[j]<-0
  ETR2<-ETMrest*RS[j]/RS0
  RS[j]<-RS[j]-ETR2
  EtatETR[Jour,j+1]<-ETR1+ETR2
  EtatStressH[Jour,j+1]<-(1-(ETR1+ETR2)/ETM)
  }
```

```
# Stockage des réservoirs
EtatRFU[Jour,j+1]<-RFU[j]
EtatRS[Jour,j+1]<-RS[j]
}
```

```
# Calculs de statistiques par agrégation de sous-échantillons
# Quantiles par paquets de (NbSimul^0,5)
# RFU
...
# RS
...
# ETR
...
# StressH
...
# Ecoule
...
}
```

```
# Ecriture des fichiers de sortie finaux
#####
```

```
# Fichiers de données
# Format R Data, compact
...
# Format texte arrondi à 2 décimales
...
# Fichiers de statistiques
# Format texte seulement
...
}
```

```
# PARTIE OBSERVATIONS #
#####
...
```

Le déroulé de la partie observations ne comprend aucune particularité.

7.2. Perspectives

Les calculs de BHR n'avaient pas été prévus au départ dans les phases du projet AP3C qui se sont achevées au 31/12/2019. Ils résultent d'un côté d'une opportunité (un créneau de temps de travail qui s'est ouvert) et de l'autre de l'ambiguïté d'interprétation de certains IAC. Par exemple un déficit pluviométrique qui s'accroît au démarrage de la saison de végétation n'est pas du tout ressenti de la même manière selon qu'on dispose d'un sol superficiel ou profond.

Ce qui a pu être fait comme traitement de l'eau dans le sol dans ces conditions a été néanmoins limité. On a juste eu le temps de produire les BHR pour un seul type de végétation (de l'herbe en permanence) et pour un panel de sols assez restreint et rigide. La seule utilisation en a été la production de statistiques basiques quotidiennes qui, en pratique, ne donnent accès qu'à des indicateurs fonctionnant en conditions moyennes.

Une première perspective consiste à avoir un panel de sols plus large, et plus adapté aux situations départementales, combiné à un panel de cultures parmi les plus répandues en situation actuelle et potentiellement vers le futur. Il faut cependant faire très attention à la masse de données produite. Ainsi, dans les conditions de AP3C, chaque calcul de BHR pour le réseau (un type de sol et une végétation) produit 600Go de données.

Or, on peut facilement arriver à une dizaine de types de sols, combinés à une centaine de cultures. Ça peut sembler excessif mais, pour ne prendre que le maïs, il faut multiplier le nombre de gammes d'indice (disons 5) par un nombre raisonnable de dates d'implantation (espacées de 10 jours par exemple, disons 5 aussi) et on arriverait déjà à 25 sortes de BHR. En tout, avec environ 1000 BHR, ça fait donc 600To de données pour un domaine de type Massif central. Quant au temps de calcul, dans les conditions matérielles de AP3C (environ 700Go de mémoire vive répartie sur des machines récentes), on arriverait à environ 20 000 heures en exécution séquentielle stricte, soit 2 000 heures avec une parallélisation en paquets de dix.

Bref, il faut réfléchir sérieusement au préalable à ce qui serait très sensible et divergent du BHR herbe pour restreindre fortement la multiplication des BHR autres.

Une deuxième perspective est un usage des données hydro-pédo-climatiques produites à l'intérieur d'indicateurs agro-climatiques de seconde génération, qu'on appellerait Indicateurs Agro-Pédo-Climatiques (IAPC). Ceci permettrait d'aborder des domaines bien plus vastes qu'avec les seules variables climatiques directes.

Avec l'écoulement, on peut penser à la recharge hivernale, que ça soit pour la gestion d'étangs ou de retenues collinaires. L'absence d'écoulement peut aboutir, via l'usage des petits cours d'eau, à la quantification de la nécessité d'abreuvement.

Avec le niveau dans la RFU, on peut aborder des questions de portance de sol ou de possibilités d'implantation de cultures après récolte.

Le niveau dans la RS ou le stress hydrique permettent d'étudier la dégradation du végétal, d'identifier les périodes d'arrêt de pousse et d'anticiper un échaudage qui ne soit pas uniquement thermique.

Enfin l'ETR, dans certaines limites (si on reste au même stade végétatif et sans destruction), permettrait d'aborder la question de l'évolution des rendements quantitatifs.

Plus généralement le BHR n'est qu'un modèle (assez simple) parmi tant d'autres. N'importe quel autre modèle (risques sanitaires, confort thermique d'un bâtiment d'élevage..) qui utilise des données météorologiques quotidiennes, peut intégrer les séries de 10 000 projections qui ont la même qualité que les séries d'observations.

CONCLUSION

Lorsqu'on survole cette documentation technique relative à la méthodologie climatique du projet AP3C, on peut être saisi de vertige de plusieurs manières : face à la masse de données nécessaires en entrée ou produites en sortie, face à un grand nombre d'étapes avec une technicité avérée, face à la chronologie de leur enchaînement, face à la maîtrise de l'environnement de travail...

Certes, ceci n'a pas la facilité d'un « tuto » destiné à apprendre les rudiments d'une nouvelle (et éphémère) application sur smartphone. Beaucoup d'entre nous en ont pris une certaine habitude mais l'usage n'est pas le même. Il ne s'agit pas que tout un chacun homogénéise des données ou fabrique des projections climatiques chez soi. Le but est de rendre possible, et beaucoup plus facile qu'au lancement du projet AP3C (autrement dit à la fin du projet pionnier équivalent sur la Creuse), une forme de répliquabilité. En particulier, il convenait d'éviter qu'elle exige impérativement une implication forte du concepteur de la méthode.

On peut entrevoir plusieurs alternatives pour aborder cet aspect, dans le cadre d'une transposition sur un autre domaine géographique par exemple.

Il est possible de partir de manière isolée en autonomie-découverte, mais dans ce cas il est suggéré qu'on y aille en groupe motivé. Il y aurait une phase de prise de connaissance et de compréhension des enjeux et nécessités de fonctionnement (en termes techniques et humains). Cela pourrait se faire de manière séquentielle (une personne = une étape) avec détermination de l'articulation entre phases successives ou alors en parallèle avec debriefing en commun à chaque fin d'étape. Ensuite, le projet concret de répliquabilité pourrait être dessiné de manière opérationnelle. Pour fortement réduire cette phase initiale, on peut également se faire assister par la structure qui a porté AP3C pour les aspects de gestion de projet et par le climatologue rédacteur de ce rapport pour tous les aspects méthodologiques, climatiques et statistiques.

Alors oui, ne cachons pas que tout ceci demandera beaucoup de moyens, de travail et même d'efforts alors que par ailleurs des projections climatiques, produites par des instituts réputés, sont à disposition gratuite via Internet. Mais, en fin de compte, vous disposerez de projections climatiques localisées, dont les trajectoires sont compatibles avec le Changement Climatique réellement engagé sur le terrain et, à l'heure où ces lignes sont écrites, ceci n'existe nulle part ailleurs. Vous souffriez, peut-être sans le savoir, de ne pas avoir des « références climatiques » utilisables opérationnellement... ça n'est pas grave, car vous allez en fabriquer !

Vous pourrez ensuite utiliser les projections climatiques pour toute problématique qui se pose à votre structure (ou aux structures de votre territoire). Il suffit de déterminer des indicateurs de suivi (par exemple les mêmes que vous utilisez sur le temps présent) et que ceux-ci soient calculables avec des données quotidiennes de précipitations, de températures, d'ETP et aussi avec les 5 variables pédo-hydro-climatiques décrites dans ce document, si vous effectuez les calculs de BHR.

Il y a quelques années de cela, on évoquait encore des mesures « sans regret » au sens où celles-ci seraient également valides en cas d'absence du Changement Climatique. Hélas, cette perspective est désormais totalement exclue. Avec l'emploi de cette méthode, vous pourrez accéder à de véritables mesures « sans regret » au sens où les trajectoires climatiques projetées correspondent, au moins pour les 30 années à venir, au minimum vraisemblable de l'évolution qui se produira réellement.