Guide d'utilisation du logiciel R *climatol* (version 4.3-0)

José A. Guijarro (jaguijarro21@gmail.com)

2025-10-14*

Table des matières

Avant-propos

Ce guide est un complément au manuel standard R inclus dans le logiciel *climatol* lui-même, où tous les aspects de chaque fonction sont détaillés (description des paramètres, informations complémentaires et exemples). Il sera expliqué ici comment utiliser ces fonctions pour le contrôle qualité, l'homogénéisation et le remplissage des données manquantes d'un ensemble de séries climatiques, et comment en obtenir des produits dérivés. Des exemples de fonctions qui génèrent divers graphiques utiles en climatologie seront également présentés, mais les informations sur son utilisation doivent être trouvées dans le manuel standard.

L'historique des modifications et des nouvelles fonctionnalités intégrées aux différentes versions est disponible (en anglais) sur NEWS.

Les mises à jour mineures (4.3-*) au fur et à mesure de leur apparition peuvent être trouvées, ainsi que ce guide de l'utilisateur et quelques vidéos d'aide, sur https://climatol.eu

Ce guide est structuré en deux chapitres. Le premier et principal est dédié à l'homogénéisation des séries et à l'interprétation des résultats, tandis que le deuxième montre des exemples d'autres fonctions.

1 Homogénéisation des séries

1.1 Introduction

Les séries d'observations météorologiques sont d'une importance capitale pour l'étude de la variabilité climatique. Cependant, ces séries sont fréquemment contaminées par des événements étrangers à ladite variabilité : erreurs dans la prise des mesures ou dans leur transmission, changements des instruments utilisés, dans la localisation de l'observatoire, ou dans son environnement. Ces derniers peuvent être des changements soudains, comme le feu d'une forêt voisine, ou graduels, comme la reprise ultérieure de la végétation. Ces altérations de la série, appelées hétérogénéités, masquent les véritables changements climatiques et font que l'étude de la série conduit à des conclusions erronées.

Pour résoudre ce problème, des méthodologies d'homogénéisation ont été développées depuis de nombreuses années pour éliminer ou réduire autant que possible ces altérations indésirables. Initialement, elles consistaient à comparer une série problème à une autre supposée homogène, mais cette hypothèse étant très risquée, des séries de référence ont été construites à partir de la moyenne d'autres sélectionnées pour leur proximité ou leur forte corrélation, diluant ainsi leurs éventuelles hétérogénéités. D'autres méthodes procèdent à la comparaison de toutes les séries disponibles deux à deux, de sorte que la détection répétée d'un changement de moyenne permet d'identifier la série erronée. Des revues de ces méthodologies peuvent être consultées dans les travaux de Peterson et al. (1998), Aguilar et al. (2003) et Venema et al. (2012), ainsi que les Directives sur l'homogénéisation (OMM, 2020).

Il y a plusiers de logiciels qui mettent en œuvre ces méthodes afin qu'elles puissent être utilisées par la communauté climatologique. L'action COST ES0601 *(Advances in homogenisation method of climate series : an integrated approach, «HOME»)* a financé un effort international pour les comparer (Venema et al., 2012). Par la suite le projet MULTITEST (Guijarro et al., 2023) a fait une autre comparaison des méthodes à jour qui pourraient être exécutées en mode entièrement automatique. Jusqu'alors, l'attention était portée sur l'homogénéisation des séries mensuelles, principalement de température et de précipitations, mais l'étude de la variabilité des événements extrêmes a suscité un intérêt croissant pour l'homogénéisation des séries journalières, et dans le projet INDECIS on homogénéisé tous les données journalières de huit variables climatiques essentielles de la base de données ECA&D.

Les résultats des comparaisons réalisées dans le cadre du projet MULTITEST précité (Guijarro *et al.*, 2023) et dans une thèse de doctorat sur l'homogénéisation des températures quotidiennes (Killick, 2016) a montré que avec l'utilisation de *climatol* la qualité des homogénéisations est parmi les meilleures que l'on puisse obtenir par d'autres méthodes. De plus, alors que quelques autres logiciels tolèrent peu l'absence de données, *climatol* été conçu pour pouvoir utiliser des séries très courtes ou fragmentées, profitant ainsi de toutes les informations climatiques disponibles dans la zone d'étude.

1.2 Méthodologie

Initialement, *climatol* était programmé pour combler les données manquantes au moyen d'estimations calculées à partir des séries les plus proches. A cet effet, la méthode de Paulhus et Kohler (1952) a été adaptée pour renseigner les précipitations journalières au moyen de moyennes de valeurs mesurées au voisinage, normalisées par division par leurs précipitations moyennes respectives. Cette méthode a été choisie pour sa simplicité et pour permettre l'utilisation de séries voisines même si elles n'ont pas de période d'observation commune avec la série problème, ce qui empêcherait le calcul des corrélations et l'ajustement des modèles de régression.

En plus de normaliser les données par division par leurs valeurs moyennes, *climatol* propose également de le faire en soustrayant les moyennes ou par standardisation complète. Ainsi, en appelant m_X et s_X la moyenne et l'écart type d'une série X, nous avons ces options pour sa normalisation :

1. Soustraire la moyenne : $x = X - m_X$
2. Diviser par la moyenne : $x = X/m_X$
3. Standardiser : $x = (X - m_X)/s_X$

Le principal problème de cette méthodologie est que les moyennes (et les écarts-types dans le troisième cas) des séries sur la période d'étude ne sont pas connues si les séries ne sont pas complètes, ce qui est le plus courant dans les bases de données réelles. Alors *climatol* calcule d'abord ces paramètres avec les données disponibles dans chaque série, complète les données manquantes à l'aide de ces moyennes et écarts-types provisoires, et les recalcule avec les séries remplies. Les données initialement manquantes sont ensuite recalculées à l'aide des nouveaux paramètres, ce qui entraîne de nouvelles moyennes et de nouveaux écarts-types, en répétant le processus jusqu'à ce qu'aucune moyenne ne change lorsqu'elle est arrondie à la précision initiale des données.

Une fois les moyennes stabilisées, toutes les données sont normalisées et on procède à leur estimation (qu'elles existent ou non, dans toutes les séries) à l'aide de l'expression simple :

$$\hat{y} = \frac{\sum_{j=1}^{j=n} w_j x_j}{\sum_{j=1}^{j=n} w_j}$$

dans laquelle $\hat{y}$ est une donnée estimée au moyen des n données correspondantes x_j les plus proches disponibles à chaque pas de temps, et w_j est le poids attribué à chacune d'entre elles.

Statistiquement, $\hat{y_i} = x_i$ est un modèle de régression linéaire appelé *axe majeur réduit* ou *régression orthogonale*, dans lequel la ligne est ajustée en minimisant les distances des points mesurés perpendiculairement à celleci (modèle II de régression) au lieu de verticalement (modèle I de régression), comme se fait généralement (figure 1), dont la formulation (avec séries normalisées) est $\hat{y_i} = r \cdot x_i$, ou r est le coefficient de corrélation entre les séries x et y. Notons que ce type d'ajustement repose sur l'hypothèse que la variable indépendante x est mesurée sans erreur (Sokal et Rohlf, 1969), hypothèse qui ne tient pas lorsque les deux sont des séries climatiques.

Les séries estimées des autres servent de références pour leurs séries observées correspondantes, de sorte que l'étape suivante consiste à obtenir des séries d'anomalies (spatiales) en soustrayant les valeurs estimées de celles observées (toujours de manière normalisée). Ces séries d'anomalies vont permettre :

— Contrôler la qualité des séries et éliminer les données dont les anomalies spatiales dépassent un seuil prédéfini `dz.max`.
— Vérifier son homogénéité en appliquant le *Standard Normal Homogeneity Test* (SNHT ; Alexandersson, 1986). Alternativement, le test de Cucconi (1968) peut être choisi, mais seulement si les séries de référence sont complètes ou presque complètes.

Lorsque les valeurs maximales obtenues lors de l'application du test à la série sont supérieures à un seuil prédéfini `inht` *(INHomogeneity Threshold)*, la série est divisée par le point de valeur maximale du test, en transmettant toutes les données précédentes à une nouvelle série qui est ajoutée à les autres avec les mêmes coordonnées mais en ajoutant un suffixe numérique au code et au nom de la station. Cette procédure est effectuée de manière itérative, en coupant uniquement les séries avec des valeurs SNHT plus élevées dans chaque cycle, jusqu'à ce qu'aucune autre inhomogénéité ne soit trouvée. De plus, le SNHT éstant un test

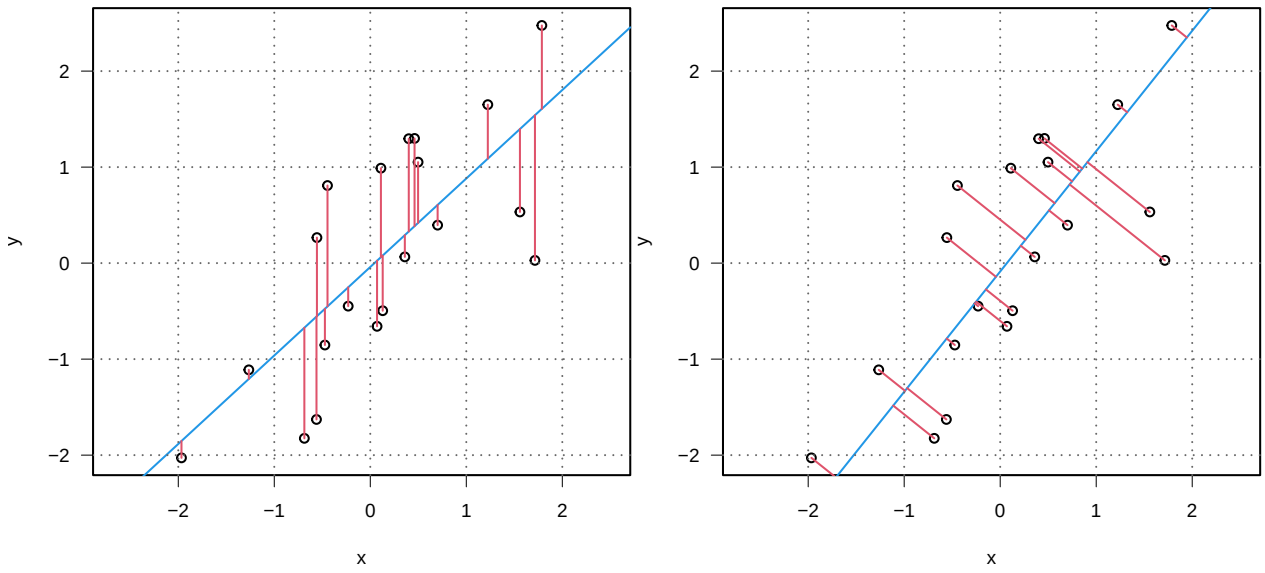


FIGURE 1 – En rouge, les écarts à la droite de régression linéaire (bleu) minimisés par les moindres carrés dans les modèles I (gauche) et II (droite).

conçu à l'origine pour trouver un seul point d'arrêt dans une série, l'existence de deux sauts dans la moyenne ou plus d'une taille similaire pourrait masquer ses résultats. Pour minimiser ce problème, dans un premier passage, SNHT est appliqué sur des fenêtres de temps qui se chevauchent, puis dans un second passage, SNHT est appliqué à les séries complétes, c'est-à-dire lorsque le test a plus de puissance de détection. Enfin, une troisième passage est dédiée au remplissage de toutes les données manquantes dans toutes les séries et sousséries homogènes avec la même procédure d'estimation des données expliquée ci-dessus. Par conséquent, bien que la méthodologie sousjacente du programme soit très simple, son fonctionnement est compliqué par une série de processus itératifs imbriqués, comme le montre l'organigramme de la figure 2.

Bien que des seuils SNHT aient été publiés pour différentes longueurs de séries et niveaux de signification statistique, l'expérience montre que ce test peut donner des valeurs très différentes selon la variable climatique étudiée, le degré de corrélation entre les séries, et leur fréquence temporelle. Climatol adopte par défaut la valeur seuil `inht=25`, adaptée aux valeurs mensuelles de température, bien qu'un peu conservatrice, en essayant de ne pas détecter de faux sauts de moyenne au prix d'en laisser passer des mineurs. Pour d'autres variables, il peut être nécessaire d'ajuster ce seuil, pour lequel les histogrammes des valeurs de test qui apparaissent dans le fichier graphique peuvent être utilisés. Si vous souhaitez homogénéiser directement des séries journalières, le seuil devrait être environ 10 fois plus grand, mais il convient de détecter les changements de moyenne sur les séries mensuelles, puis d'utiliser les points de coupure (éventuellement ajustés aux métadonnées disponibles) pour obtenir les séries journalière homogénéisées.

1.3 Procédure d'homogénéisation

Après avoir exposé la méthodologie suivie par le logiciel *climatol*, cette section sera consacrée à illustrer son application pratique à travers quelques exemples.

1.3.1 Préparation des fichiers d'entrée

Climatol n'a besoin que de deux fichiers d'entrée, un avec la liste des coordonnées, codes et noms des stations, et un autre avec toutes les données, station par station, dans le même ordre qu'elles apparaissent dans le fichier des stations. Aucun de ces fichiers n'a de ligne d'en-tête ou de numéro de ligne, et leurs données sont séparées par des espaces. (Si les noms de stations sont composés de plusieurs mots, ils doivent être placés entre guillemets). Dans le fichier de données, toutes les séries doivent être complètes, représentant les données manquantes avec `NA` ou un autre code distinctif, mais les lignes peuvent avoir n'importe quel nombre de

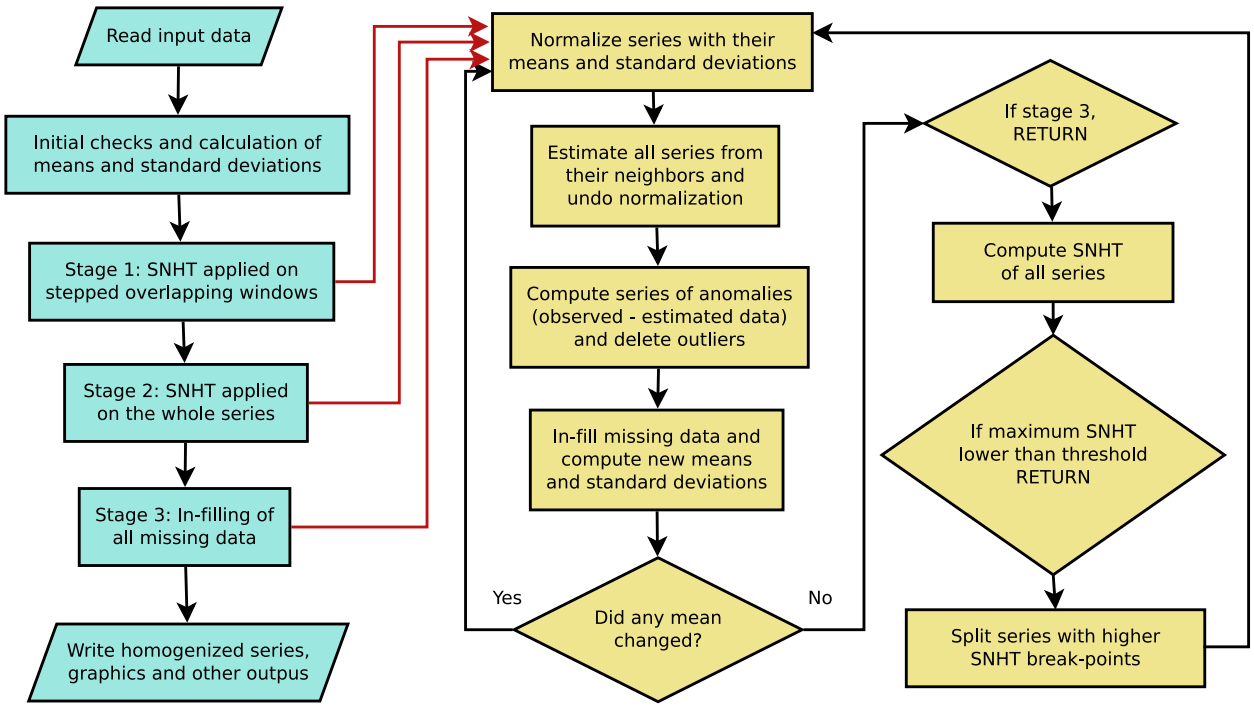


FIGURE 2 – Organigramme du fonctionnement de la fonction *homogen*, montrant ses processus itératifs.

données, car elles seront lues séquentiellement. De plus, pour que les fonctions de post-traitement décrites ci-dessous fonctionnent correctement, la période d'étude doit s'étendre sur des années complètes, commençant le janvier (le 1er s'il s'agit de données quotidiennes) de l'année initiale et se terminant le décembre (le 31 dans le cas des données quotidiennes) de la dernière année, bien que cela ne soit pas strictement nécessaire.

Les deux fichiers partagent le même nom de base `VRB_aaaa-AAAA` (où `VRB` est une abréviation de la variable étudiée, `aaaa` est la première année et `AAAA` est la dernière année couverte par les séries), mais ils ont des extensions différentes : `dat` pour les données et `est` pour les estacions. Ces extensions ne sont pas directement reconnues par Windows, les utilisateurs de ce système d'exploitation devront donc indiquer qu'elles doivent être ouvertes avec le blocnotes ou un autre éditeur de texte brut, en évitant l'utilisation de traitements de texte tels que MS-Word.

Ceci peut être illustré avec les exemples de la documentation standard du *climatol*. (Comme une exigence du référentiel CRAN est que les exemples doivent s'exécuter en quelques secondes, les données des exemples sont beaucoup plus petites que celles normalement utilisées dans les applications réelles) :

```
library(climatol) #charger les fonctions en mémoire
data(climatol_data) #charger les données d'example
#afficher un fragment du fichier de données:
write(Temp.dat[41:80,4],stdout(),ncolumns=10)
```

```
20.4 23 25.8 26.1 24.8 18.9 14.9 11.7 10.6 8.5
12.6 14.2 19.3 22.4 26.4 26.1 NA NA 15.9 12.5
13.2 NA 12.6 17 NA NA NA 25.5 23.1 18.2
12.5 10.1 10.7 11.2 13.5 NA 18.1 19.5 NA 24.9
```

```
#afficher le fichier des stations:
write.table(Temp.est,stdout(),row.names=FALSE,col.names=FALSE)
```

```
-2.5059 39.0583 210 "st01" "Station 1"
-2.7028 38.9808 112 "st02" "Station 2"
-2.63 38.8773 111 "st03" "Station 3"
-2.5699 38.9205 112 "st04" "Station 4"
-2.4663 38.9885 125 "st05" "Station 5"
```

On peut voir que chaque ligne contient les coordonnées X (longitude, °), Y (latitude, °), Z (altitude, m), le code et le nom de la station, le tout séparé par des espaces. Les coordonnées sont exprimées en degrés avec des décimales (et non en degrés, minutes et secondes) et avec le signe approprié pour indiquer l'ouest, l'est, le nord ou le sud.

Sauvons des fichiers d'entrée des températures mensuelles et des précipitations journalières pour réaliser quelques exemples d'homogénéisation (choisissez d'abord un répertoire de travail dans la session R) :

```
#températures mensuelles de 5 stations en 1961-2005:
write.table(Temp.est,'Temp_1961-2005.est',row.names=FALSE,col.names=FALSE)
write(Temp.dat,'Temp_1961-2005.dat')
#précipitations quotidiennes de 3 stations en 1981-1995:
write.table(SIstations,'Prec_1981-1995.est',row.names=FALSE,col.names=FALSE)
dat <- as.matrix(RR3st[,2:4])
#les séries de précipitations sont complètes, mais supprimons quelques données:
dat[1:300,1] <- dat[c(1000:1200,2000:2015),2] <- dat[5000:5478,3] <- NA
#et introduire également quelques erreurs:
dat[500:509,1] <- 9.9; dat[600,2] <- -9.9; dat[3000,3] <- 999
#maintenant nous les écrivons dans le fichier de données:
write(dat,'Prec_1981-1995.dat')
```

Pour aider à la préparation des fichiers d'entrée avec ce format, *climatol* fournit quelques fonctions utiles (voir la documentation standard de *climatol* pour plus de détails sur son utilisation) :

- **db2dat** crée des fichiers d'entrée directement à partir d'une base de données (accessible via le protocole ODBC).
- **daily2climatol** est utile lorsque chaque station dispose des données quotidiennes stockées dans des fichiers individuels.
- **rclimdex2climatol** génère les fichiers à partir des fichiers au format RClimDex.
- **sef2climatol** rassemble les données des fichiers SEF [1].
- **xls2csv** extrait les données des fichiers xls ou xlsx et les vide dans un seul fichier au format CSV.
- **csv2climatol** lit les données d'un seul fichier CSV et génère les fichiers pour *climatol*.

1.3.2 Homogénéisation des précipitations quotidiennes

Les précipitations quotidiennes obtenues à partir de relevés manuels de pluviomètres de type Hellmann rapportées par les observateurs laissent souvent vides les jours sans pluie, ce qui crée une ambiguïté en ne précisant pas s'il n'a pas plu ou si l'observation n'a pas pu être effectuée. Lorsque cela se produit sur un ou plusieurs jours, le relevé suivant correspond aux précipitations accumulées les jours où l'observation n'a pas été effectuée. Dans ce cas, l'observateur doit signaler l'incident en indiquant les jours où il n'a pas pu effectuer sa tâche. Ces jours peuvent alors se voir attribuer un code spécial, tel que `-1`.

Il peut également arriver qu'un manque systématique d'observations le week-end (ou certains jours) n'ait pas été correctement signalé. Si nous suspectons que cela se soit produit dans l'une de nos séries de précipitations quotidiennes, nous pouvons utiliser la fonction **weekendaccum** pour détecter les années où la fréquence des zéros sur un à trois jours consécutifs est anormalement élevée, suivie d'une fréquence excessivement basse. Si cette anomalie est détectée avec un seuil de signification prédéfini (0,01 par défaut) pour une ou plusieurs années de la série, les zéros de ces jours seront remplacés par le code d'accumulation spécifié (`-1`

1. SEF (*Station Exchange Format*) est le format utilisé dans les projets de sauvetage des données du *Copernicus Climate Change Service*.

par défaut, mais il doit correspondre au code normalement utilisé). Exemple avec les fichiers de précipitations précédemment enregistrés :

```
weekendaccum('Prec', 1981, 1995)
```

Cette commande détecte des zéros cumulatifs suspects les samedis et dimanches de 1994 à la station p064 et remplace 24 d'entre eux par le code de cumul par défaut `cumc=-1`. Le fichier de données d'origine est renommé `Prec-wkn_1981-1995.dat` au cas où l'utilisateur souhaite le récupérer pour répéter le processus, et les nouvelles séries sont écrites avec le nom d'origine `Prec_1981-1995.dat`.

Que la série originale contienne déjà des données accumulées ou que des accumulations aient été détectées avec la fonction `weekendaccum`, l'étape suivante consiste à répartir les précipitations accumulées entre les jours codés avec `cumc`, pour lesquels nous utiliserons la fonction `homogen` comme suit :

```
homogen('Prec', 1981, 1995, cumc=-1)
```

Outre la désagrégation des précipitations journalières cumulées, cette application de la fonction `homogen` effectue également un contrôle qualité initial des séries. Dans cet exemple, une donnée inférieure à zéro (hormis celles codées avec `cumc`) et une donnée excessivement élevée ont été détectées et supprimées, ainsi qu'un trop grand nombre de données identiques consécutives dans la série `p064`. Si nous n'avions pas appliqué `homogen` pour désagréger les données cumulées, il aurait été judicieux d'effectuer ce contrôle qualité initial à l'aide de la commande `homogen('Prec', 1981, 1995, onlyQC=TRUE)`, qui en plus de signaler la suppression de données excessivement anormales générerait également un fichier `Prec-QC_1981-1995.pdf`, dont les trois premiers graphiques montreraient, pour chaque série (figura 3) :

1. Boîtes à moustaches des données, indiquant les données anormales supprimées par un point rouge.
2. Boîtes à moustaches des secondes différences de la série, pour détecter les données anormales isolées.
3. Longueurs de séquences de données identiques, marquant en rouge une séquence supprimée car elle contient 10 données identiques.

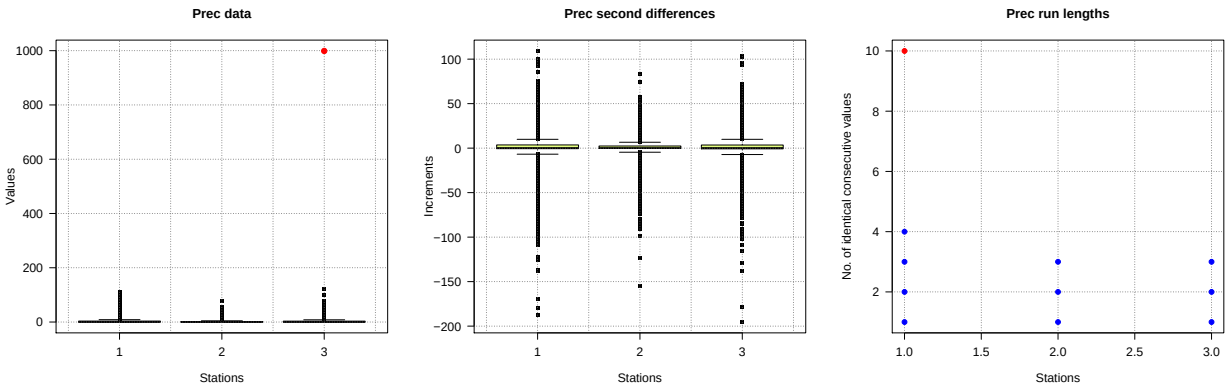


FIGURE 3 – Boîtes à moustaches et longueurs de séquences constantes pour le contrôle qualité initial.

Comme les précipitations quotidiennes ont une distribution de fréquence fortement asymétrique, la suppression des données isolées élevées est exclue car de fortes pluies peuvent survenir entre deux jours avec peu ou pas de précipitations. Avec cette variable, les zéros sont également automatiquement exclus afin de ne pas inclure ceux des jours sans précipitation dans l'analyse des séquences avec des données identiques. De cette façon, une séquence de 10 jours avec les mêmes données a été détectée dans la série 1, qui a été éliminée.

Les seuils de ce contrôle qualité initial peuvent être modifiés à l'aide du paramètre `niqd`, qui est défini par défaut sur les intervalles interquartiles `c(4,6,1)` pour les trois types de contrôle respectivement.

Après avoir trouvé des erreurs évidentes dans les données d'entrée, les fichiers originaux et les résultats du contrôle qualité sont sauvegardés en ajoutant le suffixe `-QC` au nom de la variable, et les séries exemptes desdites erreurs sont enregistrés avec les noms d'origine des fichiers d'entrée.

Nous pouvons maintenant commencer à homogénéiser les séries. La fonction qui effectue cette tâche est toujours `homogen`, qui ne nécessite initialement que la spécification des trois premiers paramètres utilisés précédemment : l'abréviation de la variable et les années de début et de fin de la période de données. Cependant, la forte variabilité des données quotidiennes (ou infra-journalières) rendant très difficile la détection des variations de la moyenne, l'homogénéisation directe de ces données n'est pas recommandée. Une erreur serait générée en recommandant d'homogénéiser d'abord les données mensuelles, ce que nous pouvons faire comme suit :

```
# valm=1 agrège les données quotidiennes en totaux mensuels au lieu de moyennes:
dd2m('Prec', 1981, 1995, valm=1) #obtenir les séries mensuelles 'Prec-m'
# nous spécifions `std=2` (normalisation recommandée pour les précipitations)
# car il ne sera pas attribué automatiquement pour les séries mensuelles.
# annual='total' fait que les derniers graphiques affichent les valeurs
# annuelles totales au lieu des moyennes:
homogen('Prec-m', 1981, 1995, std=2, annual='total')
```

Tant le fichier graphique `Prec-m_1981-1995.pdf` que celui contenant la liste des points de rupture `Prec-m_1981-1995_brk.csv` (vide dans ce cas) indiquent qu'aucune série n'a été coupée, de sorte que toutes peuvent être considérées comme homogènes. Si des sauts dans la moyenne avaient été détectés, il serait commode d'éditer les dates des points de rupture pour les ajuster aux événements de l'histoire des stations (métadonnées) qui justifient les changements détectés. Enfin, nous obtiendrons les séries journalières homogénéisées par :

```
homogen('Prec', 1981, 1995, annual='total', metad=TRUE)
```

Encore une fois, il est conseillé d'examiner les résultats pour vérifier s'il y a eu des problèmes. Un soin particulier doit être apporté à la liste des données anormales supprimées qui apparaît dans le fichier `Prec_1981-1995_out.csv`. Nous voyons qu'il y a de nombreuses données suspectes qui n'ont pas été supprimées (`Deleted=0`) et des messages de console (enregistrés dans `Prec_1981-1995.txt`) indiquent que certaines données anormales auraient été supprimées si plus d'une référence était disponible. L'histogramme des anomalies standardisées du fichier graphique `Prec_1981-1995.pdf` (figure 4) montre que deux des trois données éliminées sont assez anormales en ce qui concerne la distribution de fréquence de toutes. Cependant, la liste du fichier `Prec_1991-1995_out.csv` nous indique que les deux données correspondent au même jour (1993-10-06) dans deux séries différentes avec des anomalies de signe opposé en raison de leur divergence. Ces anomalies réciproques ont été automatiquement corrigées en marquant `Deleted=-1` dans le fichier `Prec_1981-1995_out.csv` et en appliquant la fonction `datrestore`, qui restaure les données originales dans les séries homogénéisées stockées dans le fichier binaire de résultats `Prec_1981-1995.rda`. Mais si l'une de ces données était vraiment fausse, elle devrait être supprimée du fichier d'entrée `Prec_1981-1995.dat` et l'ajustement des séries serait répété.

Cependant, il faut tenir compte du fait que même si une donnée très anormale est correcte en raison d'un phénomène météorologique local, il est préférable de l'éliminer avant l'homogénéisation et de la restituer ultérieurement dans le fichier des séries homogénéisées, car sinon cette anomalie locale entraînerait des modifications non souhaitées dans les séries voisines. (La fonction `datrestore` automatise la restauration des valeurs anormales dont l'utilisateur a manuellement changé le signe en négatif dans la colonne `Deleted` du fichier `*_out.csv`.)

1.3.3 Homogénéisation des températures mensuelles

Voyons maintenant un autre exemple d'homogénéisation avec les températures mensuelles que nous avions enregistrées :

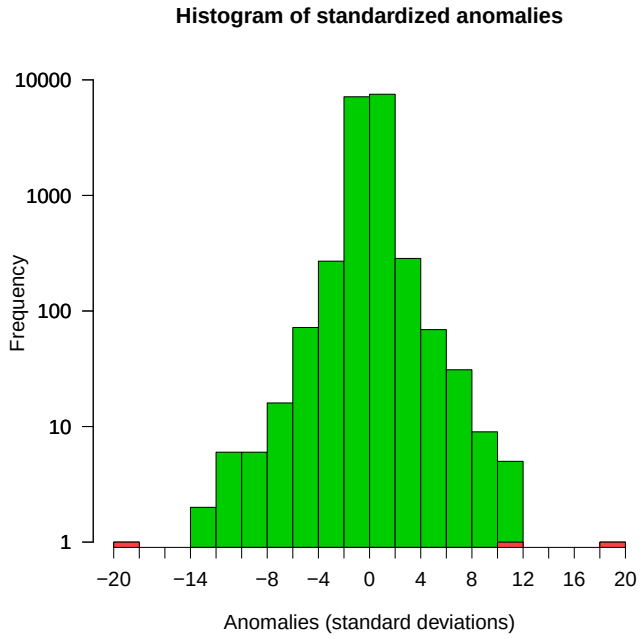


FIGURE 4 – Histogramme des anomalies spatiales montrant les données rejetées en rouge.

```
homogen('Temp', 1961, 2005)
```

Le contrôle qualité initial a supprimé une donnée hautement anormale et une série de 9 données consécutives identiques (points rouges sur la figure 5). Le fichier `Temp_1961-2005_out.csv` nous informe que cette série était composée de 9 données consécutives d'une valeur de 11 qui ont commencé le 01/02/1988.

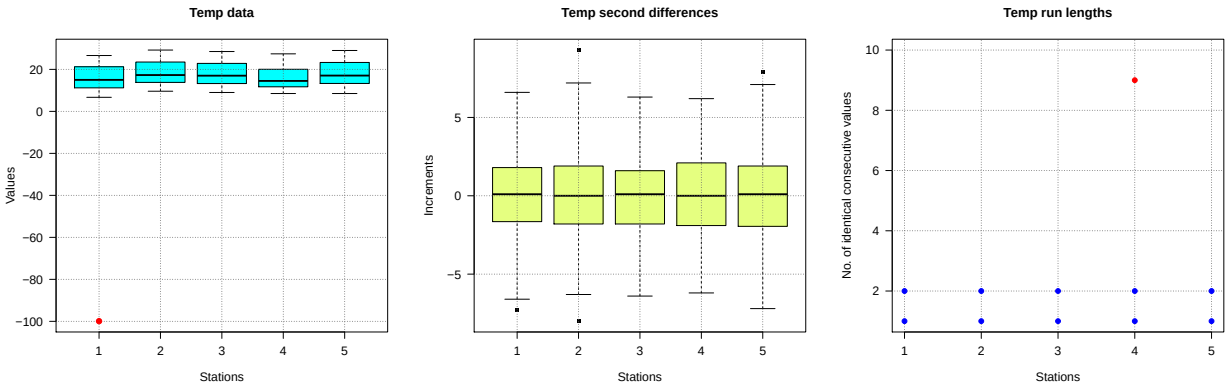


FIGURE 5 – Boîtes à moustaches et longueurs de séquences constantes pour le contrôle qualité initial.

Après les trois premiers graphiques du contrôle de qualité initial du fichier `Temp_1961-2005.pdf`, qui ne montrent que les données très anormales (automatiquement supprimées), nous pouvons voir la disponibilité des données, à la fois station par station et au total (figure 6). On voit que la série 3 est la seule série complète, alors que la série 1 est assez courte et la série 4 est très fragmentée, avec peu de données mais réparties sur toute la période d'étude. (D'autres méthodes d'homogénéisation ne pourraient pas fonctionner avec autant de données manquantes). La figure 6-droite montre combien de données existent à chaque pas de temps. Les lignes horizontales vertes et rouges en pointillés indiquent respectivement quel est le minimum souhaitable (5

données) et le minimum nécessaire (3 données) pour détecter les données suspectes, car avec seulement deux données dans un pas de temps on ne pourra pas deviner lequel sera le mauvais en cas de divergence.

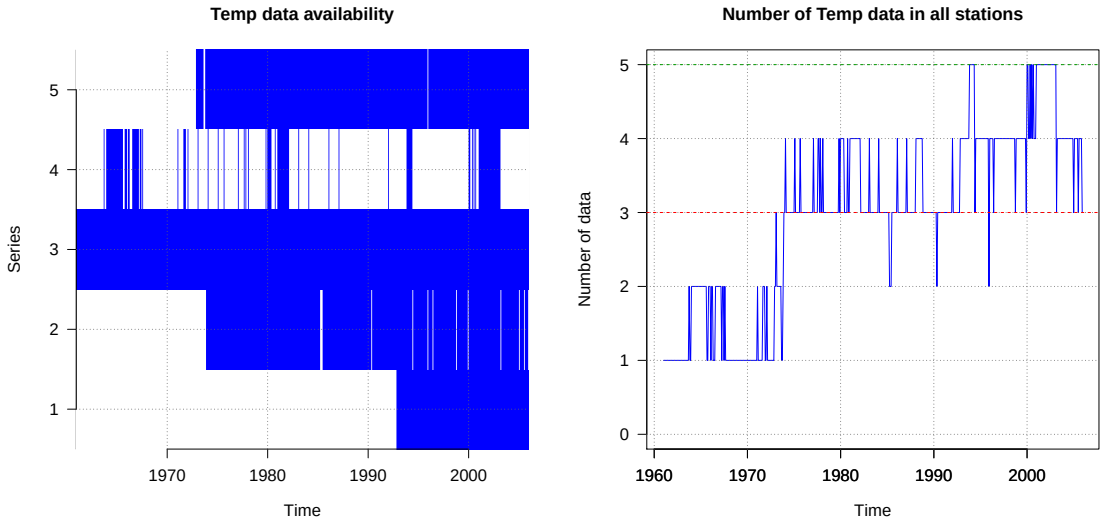


FIGURE 6 – Disponibilité des données tout au long de la période étudiée.

Le minimum absolu pour que *climatol* fonctionne est qu'il doit y avoir au moins une donnée à chaque pas de temps, car le but ultime est de combler toutes les données manquantes par interpolation spatiale, ce qui ne peut être réalisé sans données. Lorsque c'est le cas, le processus s'arrêtera avec un message d'erreur, et si cela n'est pas dû à une suppression trop importante de données anormales (ce qui peut être corrigé en abaissant le seuil `dz.max`), il faudra ajouter de nouvelles séries contenant des données dans les périodes critiques ou raccourcir la période d'étude pour éviter le problème.

Les graphiques suivants se concentrent sur les corrélations entre les séries et leur classification en groupes avec une variabilité similaire, qui sont ensuite tracées dans une carte. Les corrélations ont tendance à diminuer avec l'augmentation de la distance entre les stations. Plus les corrélations sont élevées, plus la fiabilité de l'homogénéisation et du remplissage des données manquantes est grande. En particulier, les corrélations doivent toujours être positives, au moins dans une fourchette de distances raisonnable. Sinon, il y aura probablement des discontinuités géographiques qui produisent des différences climatiques. (Par exemple, une crête de montagne peut produire différents régimes de précipitations sur ses deux côtés). Ceci peut être confirmé avec la carte des stations, dans laquelle des groupes de variabilité similaire seraient localisés dans des zones différentes, auquel cas il faudrait homogénéiser leurs séries indépendamment [2].

Dans les zones à topographie complexe et/ou à faible densité de stations, les corrélations peuvent être loin d'être optimales. Dans cette situation, les données remplies seront individuellement affectées par des erreurs majeures, mais ces paramètres statistiques devraient être acceptables.

Pour éviter de traiter des matrices de corrélation trop volumineuses, le nombre de séries utilisées pour cette analyse typologique est limité à 300 par défaut, et un échantillon aléatoire de cette taille sera utilisé lorsque le nombre de séries dépasse ce seuil, mais l'utilisateur peut le modifier via le paramètre nclust.

Après ces premiers graphiques dédiés à la vérification des données, les pages suivantes du document présentent des tracés d'anomalies (spatiales) normalisés pour chacune des trois étapes suivantes :

1. Détection dans des fenêtres étagées superposées
2. Détection en séries complètes
3. Anomalies finales des séries homogénéisées

2. La fonction `datsubset` permet d'obtenir les fichiers *climatol* d'un sousensemble de stations.

Les graphiques des deux premières étapes montrent la série d'anomalies dans lesquelles des changements de moyenne ont été détectés, marquant les points de rupture par une ligne rouge verticale en pointillés et étiquetant la valeur du test d'homogénéité à son extrémité supérieure. Les anomalies finales (dans la troisième étape) sont utilisées pour vérifier s'il y a eu des changements évidents non corrigés dans la moyenne, auquel cas il faudrait refaire l'homogénéisation en fixant un seuil d'inhomogénéité `inht` inférieur à la valeur 25 utilisée par défaut. Si, par contre, les dernières coupes dans la série d'anomalies ne semblaient pas justifiées, ce que nous ferions, c'est d'augmenter ledit seuil.

Dans notre exemple, 5 sauts de moyenne ont été détectés en quatre séries. Dans l'un d'eux, il a détecté deux sauts qui délimitaient une courte période de données anormales, qui ont été éliminées (figure 7).

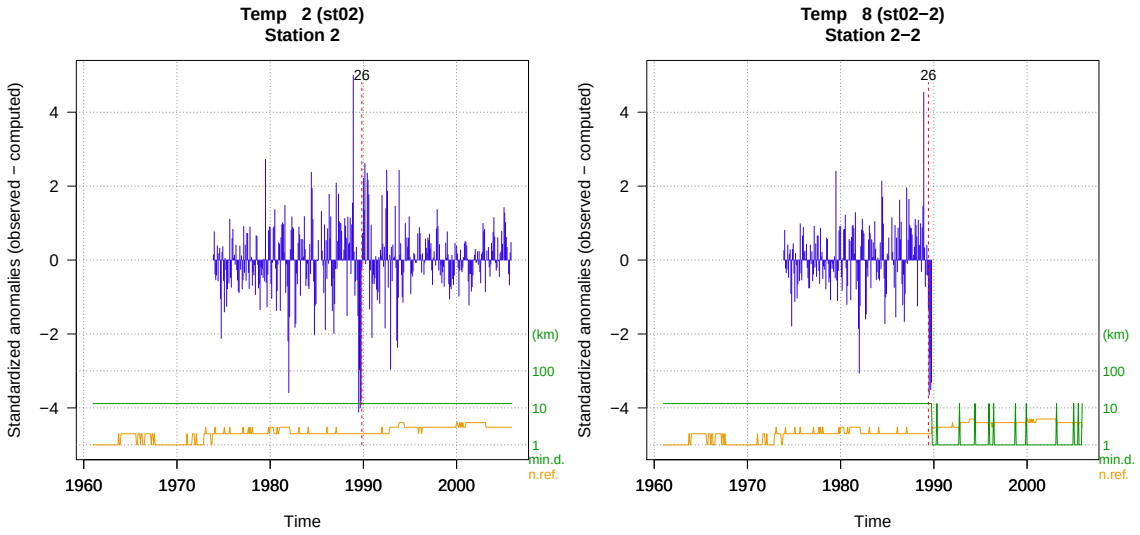


FIGURE 7 – Détection d'une courte période anormale.

Après les séries d'anomalies, nous pouvons voir les graphiques des séries reconstruites et les corrections appliquées. La figure 8 montre un exemple, avec les anomalies de la série 1 à gauche et la reconstitution de séries complètes à partir des deux fragments homogènes à droite. Le tracé des anomalies présente deux lignes supplémentaires en bas qui indiquent la distance minimale par rapport aux données voisines (en vert) et le nombre de données de référence utilisées (en orange), les deux utilisant l'échelle logarithmique de l'axe de droite. Le graphique des séries reconstruites montre leurs moyennes [3] annuelles mobiles, sauf si les séries sont très court (jusqu'à 120 termes), auquel cas toutes les valeurs sont tirées. Les séries originales sont dessinées en noir [4], et les reconstituées en couleur.

Après chaque phase du processus d'homogénéisation, des histogrammes des valeurs résiduelles du test d'homogénéité utilisé (SNHT par défaut) sont également présentés, ce qui peut aider à faire varier le seuil `inht` si le processus doit être répété. Mais quand on travaille avec peu de séries (comme dans notre exemple) il sera difficile de discriminer quelle valeur sépare le mieux les séries homogènes de celles qui ne le sont pas. Dans ce cas, il conviendra de revoir les graphiques d'anomalies, comme cela a été commenté. Comme par défaut un maximum de 4 données de référence est utilisé dans la dernière phase (au lieu du maximum de 10 utilisé dans les phases de détection), la dernière série d'anomalies peut présenter des valeurs de test supérieures au seuil utilisé, du fait de leur plus grande variabilité due à la réduction du nombre de références.

L'histogramme des anomalies qui apparaît vers la fin du document a déjà été commenté à propos de l'homogénéisation des précipitations journalières `Prec`. Enfin, la dernière page du fichier graphique contient un graphique indiquant sa qualité ou son unicité, dans lequel les stations sont localisées en fonction de leurs

3. Totaux si le paramètre `annual='total'` est donné, ce qui est recommandé pour les précipitations.
4. Mais les valeurs annuelles ne peuvent pas être calculées lorsque certaines données manquent, et elles apparaîtront alors avec la couleur du fragment reconstruit auquel ils appartiennent.

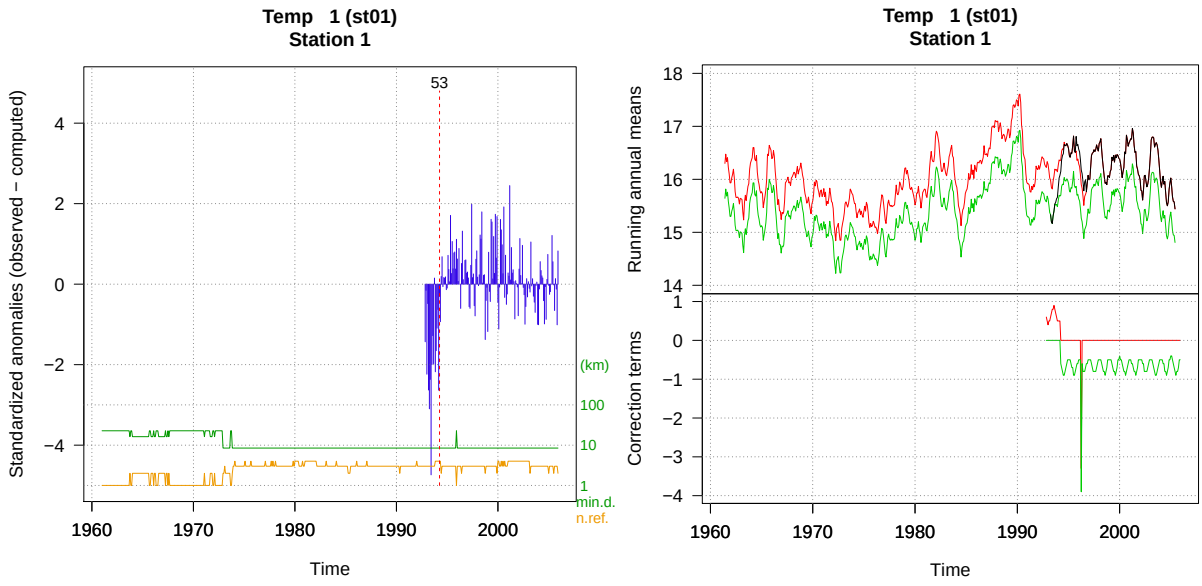


FIGURE 8 – Exemple de détection de sauts dans la moyenne (gauche) et reconstruction des séries à partir de chaque fragment homogène (droite).

erreurs types finales (RMSE) et des valeurs du test d'homogénéité. Les RMSE sont calculées en comparant les données estimées et observées dans chaque série. Une valeur élevée peut indiquer une mauvaise qualité, mais cela peut également être dû au fait que la station se trouve dans un endroit particulier avec un microclimat différent. Dans tous les cas, les séries homogènes de stations qui partagent le climat commun de la région auran tendance à se regrouper dans la partie inférieure gauche du graphique.

1.3.4 Autres paramètres de la fonction `homogen`

Cette fonction a un grand nombre de paramètres, comme on peut le voir dans sa documentation standard. En général, il n'est pas nécessaire de les préciser, puisqu'ils ont des valeurs par défaut qui sont généralement appropriées, mais selon la variable étudiée ou les premiers résultats de l'homogénéisation, il peut être opportun de modifier leurs valeurs, notamment dans les paramètres suivants :

- `dz.max` définit les seuils de rejet des données anormales ou d'avertissement concernant les données suspectes. Exemple : `dz.max=c(7,9)` supprimera les données dont l'anomalie est supérieure à 9 écarts-types, et listera comme suspectes celles qui ont des anomalies entre 7 et 9 écarts-types. Si vous souhaitez définir des seuils différents dans la queue gauche de la distribution des anomalies, des valeurs peuvent être données au paramètre `dz.min`.
- `inht` est le seuil d'inhomogénéité, c'est-à-dire la valeur du test d'homogénéité au-delà de laquelle la série sera scindée. L'examen des histogrammes de test et des graphiques d'anomalies peut suggérer de faire varier la valeur par défaut, qui est de 25. Si vous souhaitez forcer l'homogénéisation directe des séries journalières, cette valeur devra être augmentée d'un ordre de grandeur (en utilisant par exemple '`inht=250, force=TRUE`').
- `std` est le type de normalisation appliqué aux données. Si la variable est détectée comme étant fortement biaisée et bornée par zéro, `std=2` sera utilisé (la donnée sera divisée par sa valeur moyenne). Comme il s'agit de la normalisation recommandée pour les précipitations et la vitesse du vent, même si elle sera attribuée automatiquement dans les séries quotidiennes, il n'en sera pas de même pour les valeurs mensuelles, pour lesquelles il conviendra de spécifier `std=2`. La normalisation par défaut est `std=3` (soustraire la moyenne et diviser par l'écart-type), valable pour d'autres variables telles que la température, l'humidité relative, la pression atmosphérique, etc. Un troisième type de normalisation que l'utilisateur peut spécifier est `std=1`, qui ne centre les données qu'en soustrayant sa valeur moyenne.

— `vmin` et `vmax` servent à limiter les valeurs possibles que peuvent prendre les données. `vmin` est automatiquement défini sur 0 si la normalisation est `std=2`, mais par exemple pour l'humidité relative, il serait pratique de spécifier '`vmin=0, vmax=100`'.
— `nref` est le nombre maximal de données proches à utiliser pour estimer la série a tester. Par défaut, jusqu'à 10 seront utilisés (s'ils existent à chaque pas de temps considéré) dans les deux premières étapes, et jusqu'à 4 dans la dernière, mais parfois il peut être commode de changer ces valeurs. Par exemple, dans l'homogénéisation des précipitations journalières, l'utilisation de 4 données de référence va lisser les données estimées, augmentant ainsi le nombre de jours de pluie et diminuant les valeurs maximales. Ceci sera évité en fixant `nref=1`, bien qu'une valeur très élevée de la série la plus proche puisse produire une valeur trop élevée dans la série testée si la précipitation moyenne est beaucoup plus élevée que celle de la série voisine.
— `wd` spécifie la distance à laquelle le poids des données voisines est divisé par deux. Par défaut, `wd=c(0,0,100)` est défini, donc aucun poids ne sera attribué aux données dans les deux premières étapes de détection des sauts dans la moyenne et anomalies spatiales, et dans la troisième étape (remplissage final de toutes les données manquantes) les données perdront du poids avec la distance, comme le montre la figure 9, de sorte qu'à 100 km, elles pèseront la moitié.

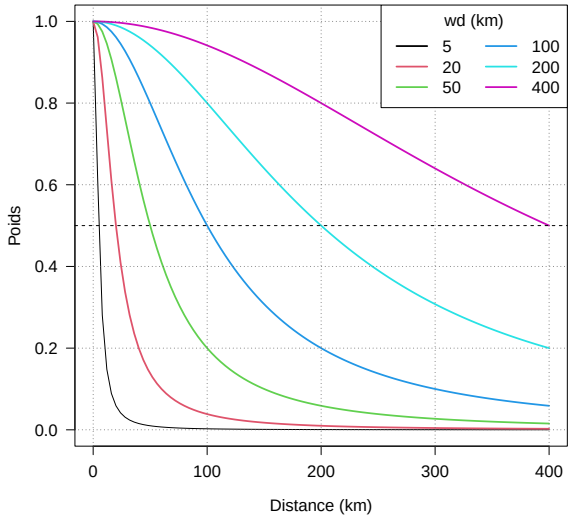


FIGURE 9 – Variation des poids pour différentes valeurs de wd.

1.3.5 Fichiers de résultats

Jusqu'à présent, les seuls fichiers non commentés que la fonction `homogen` génère également sont ceux qui stockent les résultats dans un fichier binaire R. Ils ont le même nom de base que les autres, mais avec une extension `rda`. Des détails sur son contenu sont donnés dans la documentation standard de la fonction. L'utilisateur peut charger les résultats de l'un des exemples ci-dessus dans la mémoire de travail de R pour sa manipulation au moyen de la commande :

```
load('Temp_1961-2005.rda')
```

Mais climatol fournit des fonctions de post-traitement pour faciliter l'obtention des produits des séries homogénéisées, donc dans la plupart des cas il n'est pas nécessaire d'utiliser directement les fichiers `*.rda`, comme nous le verrons plus loin.

1.4 Obtention de produits à partir des données homogénéisées

Bien que l'utilisateur puisse charger les résultats de l'homogénéisation comme indiqué ci-dessus, *climatol* fournit les fonctions de post-traitement `dahstat` et `dahgrid` pour obtenir facilement des produits fréquemment utilisés à partir des séries homogénéisées.

1.4.1 Synthèses statistiques et séries homogénéisées

Les séries homogénéisées peuvent être déversées dans deux fichiers texte CSV à l'aide de la fonction `dahstat` en spécifiant le paramètre `stat='series'`. Dans les exemples précédents, `Temp` correspondait aux températures mensuelles et `Prec` aux précipitations quotidiennes, à partir desquelles les valeurs mensuelles enregistrées étaient obtenues sous le nom de `Prec-m`. On peut donc obtenir les séries homogénéisées (ajustées du derniers fragments homogènes vers l'arrière) par :

```
dahstat('Temp', 1961, 2005, stat='series')   #températures mensuelles
dahstat('Prec', 1981, 1995, stat='series')   #précipitations quotidiennes
```

Chacune de ces commandes génère deux fichiers CSV. Les soi-disant `*_series.csv` contiennent les séries homogénéisées, tandis que les `*_flags.csv` contiennent des codes qui indiquent si les données sont observées (0), remplies (1, initialement absentes) ou corrigées (2, en raison d'inhomogénéités ou d'anomalies excessives). Avec un ordre similaire appliqué à 'Prec-m', on pourrait obtenir la série d'homogénéisation des précipitations mensuelles, mais il est préférable de calculer ces séries à partir des séries journalières, car l'absence de données journalières au moment du calcul des agrégats mensuels produira quelques différences. Pour cela vous pouvez utiliser la fonction `dahstat` avec l'option `stat='mseries'`.

Des résumés statistiques sont créés avec la même fonction. Voici quelques exemples (plus d'informations dans la documentation R de la fonction `dahstat`) :

```
dahstat('Prec',1981,1995) #moyenne mensuelle (statistique par défaut)
dahstat('Prec',1981,1995,stat='tnd') #tendances et valeurs p
dahstat('Prec',1981,1995,stat='q',prob=.2) #premier quintile
```

Cette fonction inclut des paramètres pour choisir un sous-ensemble des séries, soit en donnant une liste avec les codes souhaités, comme par exemple avec `cod=c('p064','p084')`, soit en précisant que l'on veut les séries reconstruites à partir des sous-périodes les plus longues (`long=TRUE`). On peut aussi demander les statistiques de toutes les séries (reconstruites à partir de tous les fragments homogènes) en utilisant le paramètre `all=TRUE`.

1.4.2 Séries de grilles homogénéisées

L'autre fonction de post-traitement, `dahgrid`, génère des grilles calculées à partir des séries homogénéisées (sans utiliser de données renseignées). Mais avant d'appliquer cette fonction, l'utilisateur doit définir les limites et la résolution des grilles, comme dans cet exemple en utilisant les résultats de l'homogénéisation `Temp` :

```
grd <- expand.grid(x=seq(-2.7,-2.5,.025),y=seq(38.8,39.1,.025)) #grille
#la commande suivante nécessite l'installation du package sp:
sp::coordinates(grd) <- ~x+y #convertir la grille en objet spatial
```

La fonction `expand.grid` de R est utilisée pour définir les séquences de coordonnées X et Y, puis la fonction `coordinates` (du package `sp`) est appliqué pour convertir la grille, enregistrée sous le nom grd (n'importe quel nom aurait pu être utilisé), en un objet de classe spatiale.

Des grilles homogénéisés peuvent désormais être générés (au format NetCDF) avec :

```
dahgrid('Temp', 1961, 2005, grid=grd) #grilles mensuelles
```

Ces grilles ont été construites avec des valeurs normalisées sans dimension. De nouvelles grilles avec les unités d'origine (°C dans l'exemple) peuvent être obtenues via des outils externes, tels que les *Climate Data Operators (CDO)*, en profitant du fait que `dahgrid` a également enregistré des grilles avec les moyennes (`*_m.nc`) et écarts-typiques (`*_s.nc`). Ainsi, si les `CDO` sont installés sur notre système, nous pouvons les appeler depuis R avec :

```
commande <- paste('cdo add -mul Temp_1961-2005.nc Temp_1961-2005_s.nc',
  'Temp_1961-2005_m.nc Temp-u_1961-2005.nc')
system(commande)
```

Mais les nouvelles grilles contenues dans `Temp-u_1961-2005.nc` (nous aurions pu donner n'importe quel nom au fichier de sortie, en respectant l'extension `nc`) ne seront basées que sur des interpolations géométriques, de sorte que s'il y a des montagnes dépourvues de données, les variations climatiques attendues ne seront pas reflétées dans les grilles. Pour obtenir une meilleure représentation du climat de la zone étudiée, de meilleures grilles de moyennes `Temp_1961-2005_m.nc` et d'écarts-types `Temp_1961-2005_s.nc` doivent être obtenues par des méthodes géostatistiques avant de les utiliser pour obtenir les grilles de valeurs avec ses unités d'origine.

1.5 Foire aux questions sur la fonction `homogen`

Les exemples précédents montrent et discutent des applications les plus courantes des fonctions d'homogénéisation du *climatol*. Cependant, des doutes peuvent surgir quant à la façon de procéder lorsqu'il s'agit d'autres variables climatiques ou d'autres résolutions temporelles. Cette section est dédiée à la résolution des doutes les plus fréquents.

1.5.1 Comment enregistrer les résultats de différents tests

Si vous exécutez `homogen` avec différents paramètres à explorer qui donnent de meilleurs résultats, vous pouvez éviter d'écraser les sorties précédentes en les renommant à l'aide de la fonction `outrename`. Par exemple, la commande suivante renommera tous les fichiers de sortie `Temp_1961-2005*` en `Temp-old_1961-2005*` :

```
outrename('Temp', 1961, 2005, 'old')
```

1.5.2 Comment changer le niveau de coupure dans l'analyse de clustering

Dans l'analyse de clustering effectuée par *climatol* lors de sa vérification initiale des données, le nombre de clusters est déterminé automatiquement. En regardant le dendrogramme (dans les premiers graphiques du document de sortie PDF), un niveau de coupure différent peut être choisi à l'aide du paramètre `cutlev`. Cela n'aura aucun effet sur les résultats de l'homogénéisation, car cela sert uniquement à optimiser la division des stations en groupes avec un régime climatique plus similaire au cas où l'utilisateur jugerait opportun de ne pas homogénéiser toutes les séries ensemble. (La composition des groupes apparaîtra dans la sortie texte et dans l'objet `ct` enregistré dans le fichier binaire `*.rda`).

1.5.3 Comment utiliser des séries de réanalyse comme références

Lorsque les séries sont très fragmentées et que certains pas de temps de notre période d'étude ne disposent de données dans aucune d'entre elles, ou lorsque l'on cherche à homogénéiser une série isolée, une solution est d'utiliser des séries de produits de réanalyse pour servir de références qui fournissent des données dans ces lacunes critiques.

Bien que l'apparition de nouveaux systèmes d'observation (tels que les satellites) introduit des inhomogénéités dans la quantité de données disponibles pour l'assimilation par les modèles, on peut considérer que les produits de réanalyse sont généralement plus homogènes que les séries observées. Pour utiliser ces produits comme

références, la série d'un ou plusieurs points de grille situés dans le domaine d'étude doit être ajoutée au fichier de données `*.dat`, et les coordonnées de ces points ajoutées au fichier de station `*.est`. Leurs codes doivent commencer par un astérisque (exemple : `*R43`) afin que les contrôles de qualité et d'homogénéité ignorent cettes séries plus fiables.

Comme une étude a montré que les séries de réanalyse sont de pires références que celles constituées d'observations, par défaut 1000 km de distance sont ajoutés aux séries de réanalyse afin qu'elles pèsent moins que celles observées lors de leur utilisation pour calculer les données par interpolation. Cette valeur par défaut peut être modifiée à l'aide du paramètre `raway`.

1.5.4 Quelles séries homogénéisées dois-je utiliser ?

La plupart des méthodes d'homogénéisation renvoient les séries ajustées des dernières sous-périodes homogènes, mais *climatol* génère des reconstructions complètes à partir de chaque sous-période (à moins qu'elle ne soit trop courte pour qu'une telle reconstruction soit fiable). Dès lors, l'utilisateur peut se demander lequel utiliser dans son étude climatique. La réponse dépend de l'objectif de l'enquête. Pour obtenir des valeurs normales avec lesquelles calculer les anomalies des nouvelles données entrantes pour la surveillance du climat, il utilisera les séries ajustées des dernières sous-périodes homogènes (option par défaut de `dahstat`). Mais si le but est de faire des cartes, toutes les séries doivent être considérées (en ajoutant `all=TRUE` aux paramètres `dahstat`), puis en choisissant celles qui correspondent le mieux à la variabilité spatiale de l'échelle de la carte, et en ignorant celles suspectées d'être affectées par les microclimats.

1.5.5 Le processus prend trop de temps

Cela peut arriver si nous traitons de nombreuses séries très longues avec de nombreux sauts dans la moyenne. Exemple : 400 séries thermométriques quotidiennes de 1951 à 2020. L'homogénéisation directe de ces longues séries quotidiennes peut prendre des jours, surtout si les paramètres `inht` et `dz.max` n'ont pas reçu des valeurs suffisamment élevées, car alors la série subirait un nombre élevé de coupures et rejetterait trop de données qui devront alors être remplies.

Si la procédure recommandée d'homogénéisation préalable des séries mensuelles obtenues avec la fonction `dd2m` est suivie, le temps de traitement sera beaucoup plus court. Mais dans tous les cas, il convient de se demander s'il est nécessaire d'homogénéiser toutes les séries en même temps, car il est probablement préférable de diviser nos données en ensembles plus petits, en regroupant les séries selon des sous-régions climatiquement plus uniformes. (La fonction `datsubset` peut être utilisée pour générer les fichiers pour un groupe sélectionné de stations, qui peuvent être basés sur l'analyse de regroupement générée par `homogen`, adapté par l'utilisateur en fonction de sa connaissance du climat et de la physiographie de la zone).

1.5.6 Est-ce qu'on peut utiliser *climatol* pour homogénéiser des séries de débits ?

Pour pouvoir l'utiliser, il doit exister un certain degré de corrélation (positive) entre les séries. Vous pouvez faire un test avec `onlyQC=TRUE` et voir le corrélogramme. Entre les débits provenant de différents points d'un même bassin, il est possible que la corrélation nécessaire existe (bien qu'avec probablement un certain décalage dans le temps). Une autre possibilité serait d'utiliser les débits générés avec un modèle hydrologique comme série de référence.

1.6 Références

Aguilar E, Auer I, Brunet M, Peterson TC, Wieringa J (2003) : *Guidelines on climate metadata and homogenization*. WCDMP-No. 53, WMO-TD No. 1186. World Meteorological Organization, Geneve. LINK

Alexandersson H (1986) : A homogeneity test applied to precipitation data. *Jour. of Climatol.*, 6 :661-675.

Cucconi O (1968) : Un nuovo test non parametrico per il confronto tra due gruppi campionari. *Giornale degli Economisti*, 27 :225-248.

Guijarro JA, López JA, Aguilar E, Domonkos P, Venema VKC, Sigró J, Brunet M (2023) : Homogenization of monthly series of temperature and precipitation : Benchmarking results of the MULTITEST project. Int. J. Climatol., 19 pp, DOI 10.1002/joc.8069 LINK

Khaliq MN, Ouarda TBMJ (2007) : On the critical values of the standard normal homogeneity test (SNHT). *Int. J. Climatol.*, 27 :681687.

Killick RE (2016) : Benchmarking the Performance of Homogenisation Algorithms on Daily Temperature Data. PhD Thesis, University of Exeter, 249 pp. LINK

OMM (2020) : Directives sur l'homogénéisation. OMM Nº 1245, 57 pp., Genève, Suisse, ISBN 978-92-63-21245-0. LINK

Paulhus JLH, Kohler MA (1952) : Interpolation of missing precipitation records. *Month. Weath. Rev.*, 80 :129-133.

Peterson TC, Easterling DR, Karl TR, Groisman P, Nicholls N, Plummer N, Torok S, Auer I, Böhm R, Gullett D, Vincent L, Heino R, Tuomenvirta H, Mestre O, Szentimrey T, Salinger J, Førland E, Hanssen-Bauer I, Alexandersson H, Jones P, Parker D (1998) : Homogeneity Adjustments of 'In Situ' Atmospheric Climate Data : A Review. *Int. J. Climatol.*, 18 :1493-1518.

Sokal RR, Rohlf PJ (1969) : *Introduction to Biostatistics*. 2^{nd} edition, 363 pp, W.H. Freeman, New York.

Venema V, Mestre O, Aguilar E, Auer I, Guijarro JA, Domonkos P, Vertacnik G, Szentimrey T, Stepanek P, Zahradnicek P, Viarre J, Müller-Westermeier G, Lakatos M, Williams CN, Menne M, Lindau R, Rasol D, Rustemeier E, Kolokythas K, Marinova T, Andresen L, Acquaotta F, Fratianni S, Cheval S, Klancar M, Brunetti M, Gruber C, Prohom Duran M, Likso T, Esteban P and Brandsma T (2012) : Benchmarking homogenization algorithms for monthly data. *Clim. Past*, 8 :89-115.

2 Autres fonctions

En plus des fonctions d'homogénéisation expliquées jusqu'à présent, *climatol* fournit également quelques utilitaires et produits graphiques qui seront brièvement présentés ci-dessous (voir le manuel standard pour tous les détails de son utilisation).

2.1 Utilitaires

— **`fix.sunshine`** est utilisé pour élaguer tout excès d'heures d'ensoleillement qui peut s'être produit lors de l'ajustement des séries journalières (voir exemple dans l'aide de la fonction).

— **`QCthresholds`** permet d'obtenir, pour chaque série journalière (ou sous-journalière), des quantiles mensuels de valeurs extrêmes, d'incréments entre valeurs consécutives et de séquences de valeurs identiques. Ces quantiles peuvent être utilisés pour mettre en œuvre des alertes de contrôle qualité dans les Systèmes de Gestion des Données Climatiques. Exemple pour la série `Prec` préalablement homogénéisée (une valeur minimale est fixée pour éviter de compter de longues séquences de zéros consécutifs) :

```
QCthresholds('Prec_1981-1995.rda',minval=0.1)
```

```
========== thr1: Monthly quantiles of the data
--------- Station p064
         1    2    3    4    5    6    7     8    9    10   11   12
0      0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0  0.0   0.0  0.0  0.0
0.001  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0  0.0   0.0  0.0  0.0
0.01   0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0  0.0   0.0  0.0  0.0
0.99  35.1 40.9 44.5 33.1 31.6 31.0 31.9  49.5 46.7  60.2 63.9 42.9
0.999 54.1 50.8 56.5 54.2 69.4 44.6 48.3  91.2 71.7 120.2 91.9 56.1
1     54.3 53.8 57.1 62.7 92.8 49.0 54.2 110.1 73.1 140.0 93.8 57.1
--------- Station p084
         1    2    3    4    5    6    7    8     9    10   11   12
0      0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0  0.0
```

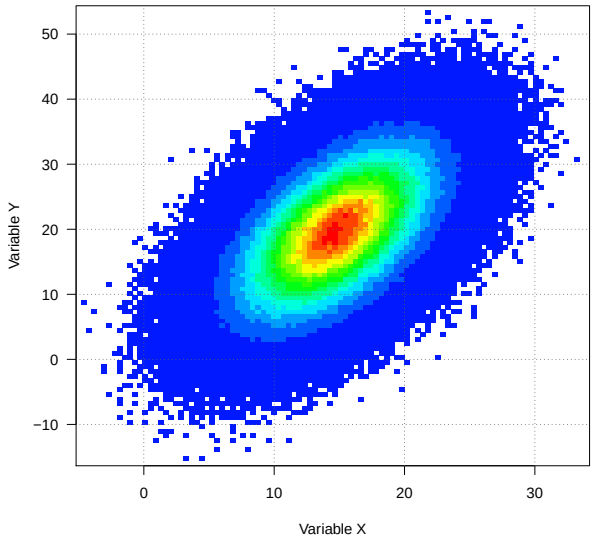
```
0.001  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0  0.0
0.01   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0   0.0   0.0  0.0  0.0
0.99  24.1 31.3 29.0 27.9 30.0 29.7 30.9 36.6  38.0  51.5 49.2 30.1
0.999 39.8 42.2 40.7 46.3 50.1 46.5 40.4 67.5  74.6  87.8 67.2 39.8
1     42.2 48.2 44.2 52.8 51.1 51.6 42.0 76.7 100.4 116.5 71.3 41.1
--------- Station p082
         1    2    3    4    5    6    7    8    9   10   11   12
0      0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.001  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.01   0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.99  17.9 24.2 21.4 23.1 21.0 26.4 28.2 33.2 29.3 38.2 31.8 19.7
0.999 24.3 32.7 29.9 31.5 34.4 34.0 47.1 48.0 61.4 63.4 48.0 24.5
1     26.9 37.6 30.4 33.8 36.8 34.0 47.9 52.0 78.7 78.0 51.2 26.3

========== thr2: Quantiles of the first differences
     0.99 0.999     1
p064 46.1  84.0 129.4
p084 37.8  66.9 103.6
p082 28.8  49.2  78.5

========== thr3: Quantiles of run lengths of constant values >= 0.1
     0.99 0.999 1
p064    2     3 4
p084    2     3 3
p082    2     3 3

Thresholds thr1,thr2,thr3 saved into QCthresholds.Rdat
(Rename this file to avoid overwriting it in the next run.)
```

Avec `load('QCthresholds.Rdat')` nous aurions ces résultats dans la mémoire de session R et nous pourrions les écrire dans le format approprié pour les importer dans le Système de Gestion des Données Climatiques afin de mettre en place des alertes de valeurs suspectes.

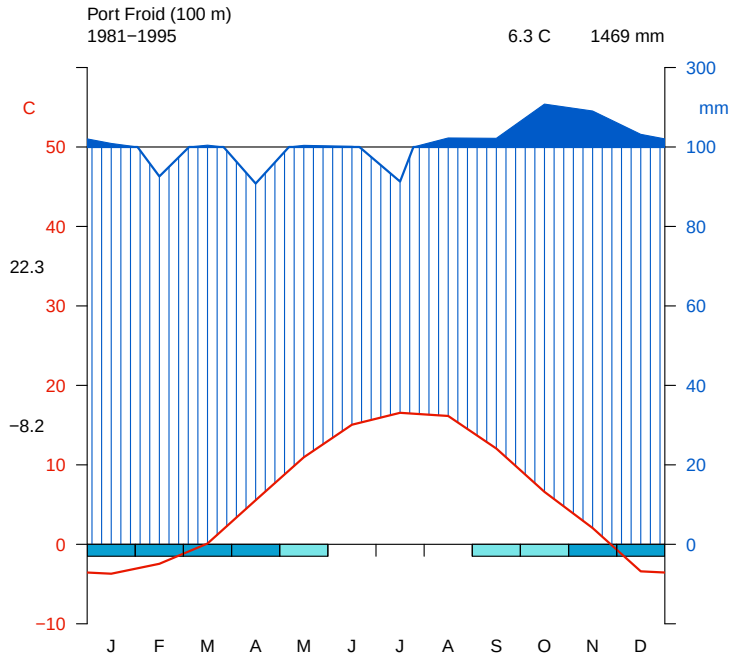
2.2 Produits graphiques

Dans cette section, seuls des exemples de fonctions produisant des graphiques utiles en climatologie seront présentés. La documentation standard de *climatol* donne le détail de chacun d'eux.

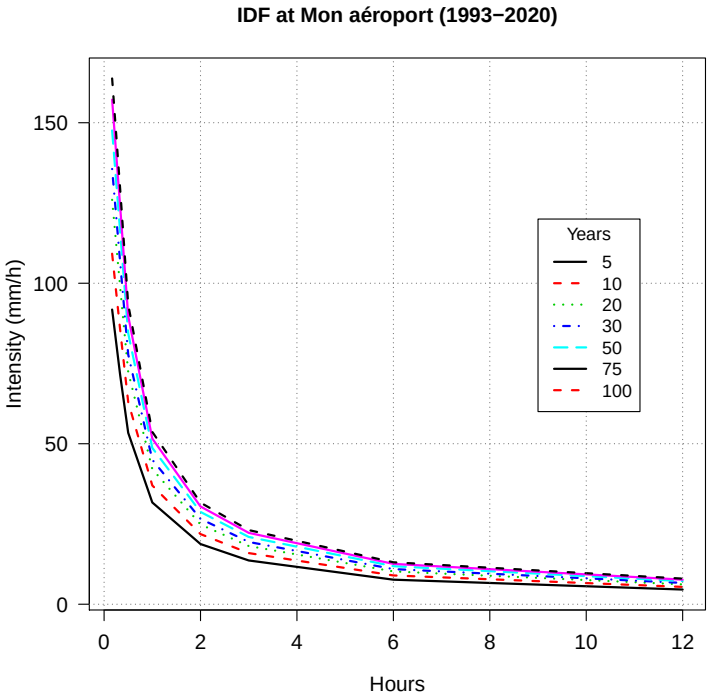
2.2.1 dens2Dplot : Nuage de points bidimensionnel



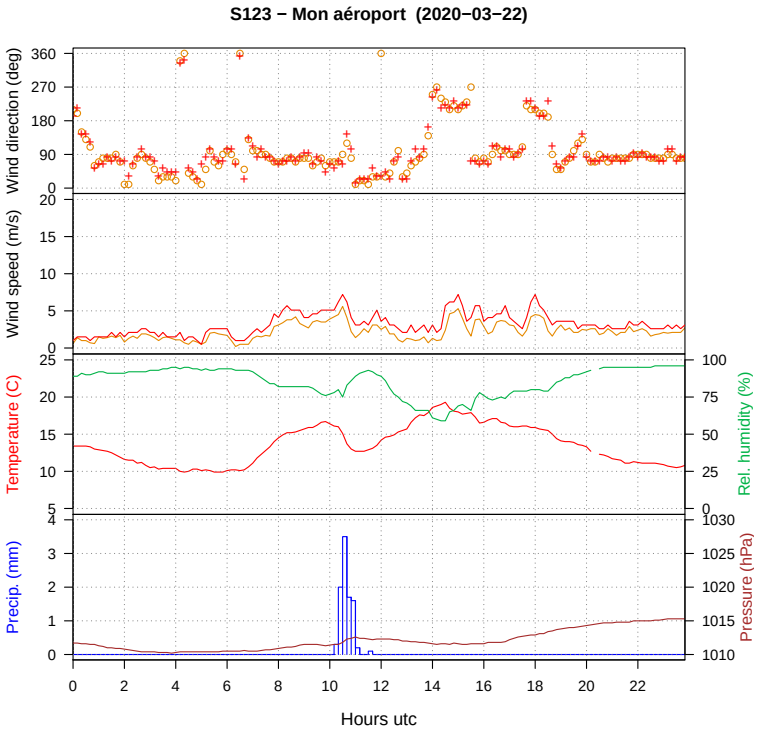
2.2.2 diagwl : Diagramme de Walter & Lieth



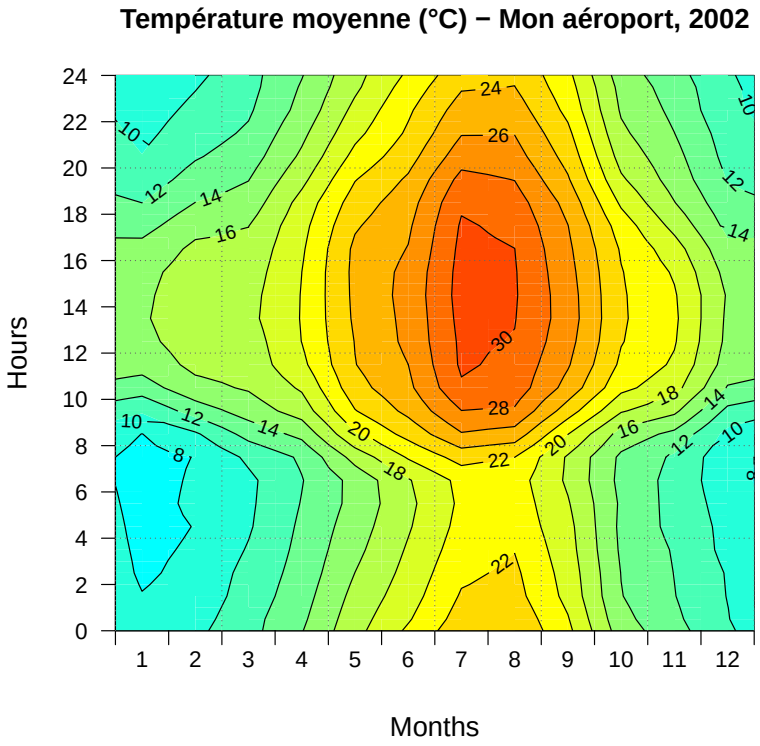
2.2.3 IDFcurves : Diagramme Intensité-Durée-Fréquence à partir des données de précipitations sous-journalières



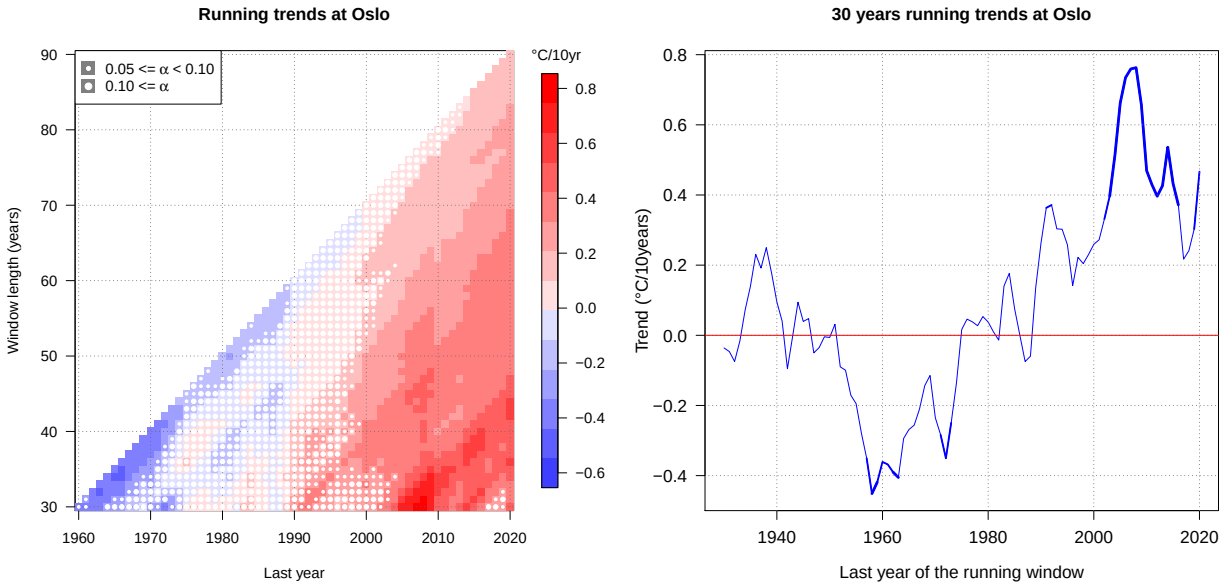
2.2.4 meteogram : Météogramme pour 1 jour



2.2.5 MHisopleths : Isoplèthes dans un diagramme Mois-Heures



2.2.6 runtnd : Diagrammes de tendances mobiles



2.2.7 windrose : Rose des vents à partir des données de direction et vitesse du vent