

# Directives sur l'homogénéisation

Édition 2020

TEMPS CLIMAT EAU



ORGANISATION  
MÉTÉOROLOGIQUE  
MONDIALE

OMM-N° 1245



# Directives sur l'homogénéisation

Édition 2020



ORGANISATION  
MÉTÉOROLOGIQUE  
MONDIALE

OMM-N° 1245

#### NOTE DE L'ÉDITEUR

La base de données terminologique de l'OMM, METEOTERM, peut être consultée à l'adresse <https://public.wmo.int/fr/meteoterm>.

Il convient d'informer le lecteur que lorsqu'il copie un hyperlien en le sélectionnant dans le texte, des espaces peuvent apparaître après <http://>, <https://>, <ftp://>, <mailto:>, et après les barres obliques (/), les tirets (-), les points (.) et les séquences de caractères (lettres et chiffres). Il faut supprimer ces espaces de l'URL ainsi recopiée. L'URL correcte apparaît lorsque l'on place le curseur sur le lien. On peut aussi cliquer sur le lien et copier l'adresse qui s'affiche dans le ruban du navigateur.

OMM-N° 1245

© **Organisation météorologique mondiale, 2020**

L'OMM se réserve le droit de publication en version imprimée ou électronique ou sous toute autre forme et dans n'importe quelle langue. De courts extraits des publications de l'OMM peuvent être reproduits sans autorisation, pour autant que la source complète soit clairement indiquée. La correspondance relative au contenu rédactionnel et les demandes de publication, reproduction ou traduction partielle ou totale de la présente publication doivent être adressées au:

Président du Comité des publications  
Organisation météorologique mondiale (OMM)  
7 bis, avenue de la Paix  
Case postale 2300  
CH-1211 Genève 2, Suisse

Tél.: +41 (0) 22 730 84 03  
Fax: +41 (0) 22 730 81 17  
Courriel: [publications@wmo.int](mailto:publications@wmo.int)

ISBN 978-92-63-21245-0

NOTE

Les appellations employées dans les publications de l'OMM et la présentation des données qui y figurent n'impliquent, de la part de l'Organisation météorologique mondiale, aucune prise de position quant au statut juridique des pays, territoires, villes ou zones, ou de leurs autorités, ni quant au tracé de leurs frontières ou limites.

La mention de certaines sociétés ou de certains produits ne signifie pas que l'OMM les cautionne ou les recommande de préférence à d'autres sociétés ou produits de nature similaire dont il n'est pas fait mention ou qui ne font l'objet d'aucune publicité.

# TABLE DES MATIÈRES

|  | <i>Page</i> |
|--|-------------|
| REMERCIEMENTS .....  | v           |
| INTRODUCTION .....   | 1           |
| <b>CHAPITRE 1. CONDITIONS PRÉALABLES .....</b>   | <b>3</b>    |
| 1.1 Collaboration avec les responsables de réseaux de stations d'observation ....                | 4           |
| 1.2 Mesures parallèles .....   | 4           |
| 1.3 Sauvetage des données .....  | 5           |
| 1.4 Contrôle de la qualité .....   | 5           |
| 1.5 Historique et métadonnées des stations .....   | 6           |
| 1.5.1 Que sont les métadonnées ? .....   | 6           |
| 1.5.2 Valeur et limites des métadonnées .....  | 7           |
| 1.5.3 Identifiants des stations .....  | 8           |
| 1.5.4 Formats et accessibilité des métadonnées .....   | 8           |
| 1.6 Formation .....  | 9           |
| <b>CHAPITRE 2. L'HOMOGENÉISATION DANS LA PRATIQUE .....</b>                                      | <b>10</b>   |
| 2.1 Homogénéisation physique et statistique .....  | 11          |
| 2.2 Sélection des données à homogénéiser .....   | 12          |
| 2.2.1 Quelles stations convient-il de sélectionner? .....  | 12          |
| 2.2.2 Comment préparer les données pour l'homogénéisation .....                                  | 13          |
| 2.3 Détection statistique ou détermination des points de changement .....                        | 14          |
| 2.3.1 Intégration des métadonnées dans les tests statistiques .....                              | 15          |
| 2.4 Séries de référence .....  | 17          |
| 2.4.1 Chevauchement .....  | 19          |
| 2.4.2 Réduction du bruit .....   | 19          |
| 2.4.3 Référence contenant des inhomogénéités .....   | 20          |
| 2.4.4 Références avec des signaux climatiques similaires .....                                   | 20          |
| 2.4.5 Utilisation des références .....   | 21          |
| 2.5 Solutions possibles en l'absence de stations de référence utiles .....                       | 21          |
| 2.6 Ajustements statistiques .....   | 23          |
| 2.7 Examen des données et cycles d'homogénéisation multiples .....                               | 25          |
| 2.8 Documentation .....  | 27          |
| 2.9 Maintenance opérationnelle d'un jeu de données homogénéisé .....                             | 27          |
| 2.10 Problèmes à l'échelle du réseau et solutions pour les résoudre .....                        | 28          |
| 2.11 Défis spécifiques aux jeux de données multinationaux .....                                  | 29          |
| 2.12 Conclusion: bonnes pratiques en matière d'homogénéisation .....                             | 30          |
| <b>CHAPITRE 3. CHOISIR UN LOGICIEL D'HOMOGENÉISATION STATISTIQUE .....</b>                       | <b>32</b>   |
| 3.1 Méthodes d'homogénéisation statistique .....   | 32          |
| 3.2 Efficacité des méthodes d'homogénéisation statistique .....                                  | 36          |
| 3.2.1 Principes théoriques .....   | 36          |
| 3.2.2 Études numériques .....  | 37          |
| 3.3 Méthodes automatisées et manuelles .....   | 39          |
| 3.4 Cas d'utilisation .....  | 39          |
| <b>CHAPITRE 4. HISTORIQUE DE L'HOMOGENÉISATION .....</b>   | <b>41</b>   |
| 4.1 Détection et ajustement .....  | 41          |
| 4.2 Séries de référence .....  | 41          |
| 4.3 Développements récents .....   | 42          |
| <b>CHAPITRE 5. CONTEXTE THÉORIQUE DE L'HOMOGENÉISATION .....</b>                                 | <b>43</b>   |
| 5.1 Structure générale des modèles spatio-temporels additifs .....                               | 43          |
| 5.2 Méthode de comparaison des séries dans le cas d'un modèle additif .....                      | 45          |
| 5.3 Méthode de détection du point de rupture (point de changement) .....                         | 46          |
| 5.3.1 Détection des points de rupture fondée sur l'estimation de<br>vraisemblance maximale ..... | 46          |
| 5.3.2 Détection des points de rupture fondée sur les tests d'hypothèses ..                       | 47          |

|   | <i>Page</i> |
|---|-------------|
| 5.3.3 Attribution des points de rupture détectés pour les séries à traiter . .    | 47          |
| 5.4 Méthodes applicables pour l'ajustement des séries . . . . .                   | 47          |
| 5.5 Options applicables pour l'évaluation et la validation des méthodes . . . . . | 48          |
| <b>GLOSSAIRE . . . . .</b>  | <b>49</b>   |
| <b>RÉFÉRENCES . . . . .</b>   | <b>51</b>   |
| <b>AUTRES LECTURES . . . . .</b>  | <b>56</b>   |

## REMERCIEMENTS

Les présentes directives sont une mise à jour de la publication intitulée *Guidelines on Climate Metadata and Homogenization* (Directives sur les métadonnées climatiques et l'homogénéisation) (WMO/TD-No. 1186). Elles ont été élaborées par l'Équipe spéciale chargée de l'homogénéisation, qui relève de la Commission de climatologie.

Nous remercions les personnes suivantes pour leur contribution exceptionnelle à la présente publication:

|                                  |  |
|----------------------------------|--|
| Victor Venema<br>(coordonnateur) | Institut météorologique, Université de Bonn (Allemagne)  |
| Blair Trewin                     | Service météorologique australien, Melbourne (Australie)   |
| Xiaolan L. Wang                  | Division de la recherche climatique, Direction générale des sciences et de la technologie, Environnement et Changement climatique Canada, Toronto, Ontario (Canada)                                  |
| Tamás Szentimrey                 | Varimax Limited Partnership, Budapest (Hongrie)  |
| Monika Lakatos                   | Service météorologique hongrois, Division du climat (Hongrie)  |
| Enric Aguilar                    | Université Rovira i Virgili, Tarragone (Espagne)   |
| Ingeborg Auer                    | Zentralanstalt für Meteorologie und Geodynamik (ZAMG), Vienne (Autriche)   |
| Jose A. Guijarro                 | Service météorologique national (AEMET) (Espagne)  |
| Matthew Menne                    | Section sur les jeux de données, Centres nationaux d'information sur l'environnement de l'Administration américaine pour les océans et l'atmosphère (NOAA), Asheville, Caroline du Nord (États-Unis) |
| Clara Oria                       | Servicio Nacional de Meteorología e Hidrología (SENAMHI) (Pérou)   |
| Wilfrid S. R. L.<br>Louamba      | Direction de la météorologie, Brazzaville (Congo)  |
| Ghulam Rasul                     | Service météorologique national, Islamabad (Pakistan)  |

Les personnes suivantes ont également apporté leur contribution:

|                      |          |
|----------------------|----------|
| Athanassios Argiriou | Grèce    |
| Peer Hechler         | OMM      |
| Gregor Vertačnik     | Slovénie |
| Yizhak Yosef         | Israël   |

---





## INTRODUCTION

Il est essentiel de disposer de séries chronologiques de données qui soient homogènes et de haute qualité pour analyser la variabilité du climat et le changement climatique.

L'homogénéisation vise à rendre les données «homogènes». Ce mot vient du grec ancien et signifie «de même nature». Malheureusement, la plupart des séries chronologiques de données climatologiques brutes portant sur de longues périodes ne répondent pas à ce principe et ne sont pas homogènes en elles-mêmes, c'est-à-dire qu'elles ne sont pas d'une seule et même nature et ne se prêtent donc pas à une analyse statistique du changement climatique. Des facteurs non climatiques, tels que de nouvelles conditions de mesure, peuvent avoir des répercussions considérables sur les valeurs mesurées et provoquer des distorsions dans le comportement statistique des séries chronologiques. L'impact de ces distorsions peut être comparable à celui du changement climatique et conduire à des conclusions erronées.

Chaque valeur qui s'affiche sur l'écran d'ordinateur d'un utilisateur résulte d'un ensemble de procédures consécutives: un jeu d'instruments est déployé en un lieu, ensuite la valeur est mesurée, enregistrée, subit des transformations et est insérée dans une base de données, puis elle est transmise aux utilisateurs intermédiaires et finals. En ce qui concerne les séries chronologiques portant sur de longues périodes, certaines de ces procédures peuvent avoir été modifiées au fil des ans. Pour une série chronologique de températures commençant à la fin du XIX<sup>e</sup> siècle, le thermomètre a très probablement été changé plusieurs fois, voire remplacé par un capteur électronique au cours des dernières décennies. Quant à l'abri, on est probablement passé d'un abri ouvert à un abri Stevenson, puis éventuellement à un abri à ailettes si la station a été automatisée. Vers le milieu du XX<sup>e</sup> siècle, de nombreuses stations d'observation ont été déplacées et installées dans les aéroports pour répondre à l'accroissement de la demande de la part de l'aviation civile. Si la station est restée au même endroit, les environs ont probablement changés. Par exemple, si la station a été déployée en périphérie d'un village il y a 100 ans, elle est peut-être entourée de bâtiments aujourd'hui. Peut-être qu'en raison de changements dans l'utilisation des sols ou pour d'autres raisons pratiques, la station a été déplacée vers un endroit plus commode en banlieue. Si nous combinions les observations faites à des périodes aussi différentes, elles ne seraient évidemment pas comparables. Il est rare, mais pas impossible, qu'un relevé de données climatologiques de 100 ans ou plus soit vraiment homogène. Par exemple, s'agissant du jeu de données australien sur les températures ACORN-SAT, seules les données de 2 des 112 stations se sont avérées homogènes pour ce qui est des températures maximales et minimales (ces 2 stations ont toutes deux commencé à fonctionner dans les années 1940). En Europe, on estime que la période entre deux ruptures détectées dans les observations de température est, en moyenne, d'environ 20 ans.

Il s'ensuit que nous ne pourrions pas tirer de conclusions définitives sur l'évolution temporelle des séries chronologiques de données climatologiques (par exemple, pour calculer des tendances fiables) sans nous assurer que toutes les observations et les séries chronologiques qui en découlent sont comparables. Aucune série chronologique de données climatologiques ne devrait être utilisée sans qu'il n'ait été procédé à des tests d'homogénéisation et à des ajustements, le cas échéant, et tous les Services météorologiques et hydrologiques nationaux (SMHN) et les fournisseurs de données climatologiques qui établissent et communiquent des jeux de données climatologiques devraient procéder régulièrement à leur homogénéisation.

Il existe deux types fondamentaux d'homogénéisation: l'homogénéisation des *moyennes* annuelles, saisonnières ou mensuelles, et l'homogénéisation de la *distribution*, qui permet d'ajuster également la variabilité autour de la moyenne ainsi que les statistiques d'ordre supérieur. Les présentes directives portent sur l'homogénéisation des moyennes.

Les jeux de données climatologiques, quels qu'ils soient, ne sont généralement pas homogènes. Les présentes directives se limitent toutefois à l'homogénéisation des données instrumentales émanant de stations terrestres.

De nombreux enseignements seront applicables à l'homogénéisation des données de radiosondage (Jovanovic *et al.*, 2017, Haimberger *et al.*, 2012) et d'autres types de données. Pour plus d'informations sur l'homogénéisation des données maritimes *in situ*, voir Kent *et al.* (2016), Kennedy *et al.* (2011) et Huang *et al.* (2015). Certains articles sur les distortions et l'homogénéisation des données satellitaires sont de Schröder *et al.* (2016) et Brogniez *et al.* (2016).

Les chapitres 1 à 3 des présentes directives visent à guider les premiers pas dans le domaine de l'homogénéisation, tandis que les chapitres 4 et 5 abordent des sujets plus pointus et plus généraux destinés aux utilisateurs expérimentés et aux développeurs de méthodes d'homogénéisation. Les explications plus détaillées sont imprimées dans une police plus petite.

Au moment de la rédaction du présent document, une page de foire aux questions (FAQ) était disponible à l'adresse suivante: <https://homogenisation.grassroots.is/>. Ces questions peuvent être de nature plus pratique ou plus temporaire (par exemple, les bugs informatiques et leur solution).

---

## CHAPITRE 1. CONDITIONS PRÉALABLES

L'homogénéisation est une étape dans le traitement des données climatiques (voir la figure 1). Les étapes précédentes, telles que le sauvetage des données et le contrôle de la qualité, ont des répercussions sur la qualité de l'homogénéisation.



Figure 1. Traitement des données climatiques

Le sauvetage des données est particulièrement important pour les régions et les périodes où les données sont rares, y compris dans les cas où nous disposons bien de données mais où la densité des stations n'est pas suffisante pour mettre au jour des problèmes de données importants. La mesure dans laquelle les données ont pu être homogénéisées devrait être prise en compte dans l'analyse ultérieure des données climatiques.

Avant d'homogénéiser un jeu de données, il est important de savoir comment la variable a été mesurée par le passé dans l'ensemble du réseau d'observation et quelle a été la situation des stations. Au cours du processus d'homogénéisation, il est essentiel de prendre conscience de la quantité de données manquantes. Si le volume de données manquantes dépasse un seuil au cours d'une période donnée, certaines méthodes d'homogénéisation

peuvent ne pas donner les résultats attendus. Les métadonnées sont également importantes lors de l'homogénéisation. Elles servent à valider les résultats de l'homogénéisation statistique et renseignent sur ce qui s'est passé dans des situations compliquées.

La dernière section du présent chapitre souligne l'importance que revêt la formation et cite quelques réunions scientifiques qui ont eu lieu récemment sur l'homogénéisation.

### 1.1 **Collaboration avec les responsables de réseaux de stations d'observation**

Il est plus simple de constituer et d'entretenir des jeux de données de séries chronologiques qui soient de haute qualité lorsqu'il y a moins d'inhomogénéités à prendre en compte. Bien que rien ne puisse être fait pour prévenir les ruptures d'homogénéité qui sont déjà présentes, il est très utile de gérer un réseau de stations d'observation de manière à réduire le nombre d'inhomogénéités dans les données actuelles et futures, et/ou à faciliter leur quantification précise. Il est nécessaire à cette fin que les climatologues collaborent avec les responsables de réseaux de stations, ce qui peut être difficile, surtout si la gestion du réseau d'observation et celle des données climatologiques sont réparties entre différents organismes.

Les méthodes ci-dessous peuvent contribuer à limiter les inhomogénéités récurrentes:

- Veiller à ce que tout nouveau système d'observation fasse l'objet d'essais sur le terrain avant d'être déployé dans l'ensemble du réseau de stations. Ces essais devraient être réalisés de préférence pendant au moins deux ans (des résultats trompeurs peuvent être obtenus sur une seule année si celle-ci est inhabituelle du point de vue climatique) et ils devraient se dérouler, en particulier dans les grands pays, dans des climats représentatifs de l'ensemble des climats du pays;
- Veiller à ce que des observations soient effectuées en parallèle en cas de changement significatif, qu'elles commencent suffisamment tôt pour fournir au moins deux années de données qui se chevauchent et que lesdites données soient archivées et partagées (voir la section 1.2 ci-dessous);
- Recenser très tôt les sites qui risquent de fermer ou dont l'environnement risque de subir des modifications importantes (par exemple, en raison de travaux de construction prévus à proximité), afin de maximiser le temps disponible pour effectuer des observations en parallèle avec tout nouveau site;
- Lorsqu'un nouveau site est choisi pour remplacer un site en cours d'exploitation, sa topographie et son environnement devrait correspondre le plus étroitement possible à ceux du site actuel, afin d'atténuer le risque de changement important sur le plan climatologique;
- Sélectionner, si possible, des sites dont il est peu probable que l'environnement change au cours de la ou des décennie(s) à venir, ou des siècles futurs;
- Le recueil permanent de métadonnées est également important. Les identifiants des stations ne devraient pas être réutilisés, car cela peut créer une confusion et conduire à la fusion de données provenant de deux stations sans aucun lien entre elles. La réutilisation des identifiants de stations est un problème, en particulier pour les jeux de données internationaux.

Un rapport de l'OMM utile sur ce point, intitulé *Protocol Measurement Infrastructure Changes* (Brandsma *et al.*, 2019), contient un exemple de protocole pour la gestion des changements dans les réseaux d'observation climatologique.

### 1.2 **Mesures parallèles**

Pour ce qui est des changements majeurs dans les observations, il est recommandé de mener des observations parallèles entre deux systèmes. La publication OMM (2018a) suggère qu'il faudrait comparer durant une longue période (un an au moins; voir le *Guide des pratiques climatologiques*

[...] les observations exécutées avec les nouveaux instruments à celles provenant de l'ancien système de mesure avant d'abandonner ce dernier». La version actualisée du *Guide des pratiques climatologiques* (OMM, 2018b) recommande que «dans la mesure où cela est réalisable, des observations simultanées devraient être effectuées – sur l'ancien et le nouvel emplacement ou avec les anciens et les nouveaux instruments – pendant une période de chevauchement d'au moins une année, et de préférence de deux ou plusieurs années, afin de déterminer les effets des modifications apportées aux instruments ou à l'emplacement des stations sur les données climatologiques». Cette fourchette temporelle suggérée montre que plus la période de chevauchement est longue, mieux c'est, mais que les période plus courtes ne sont pas à rejeter pour autant car cela vaut mieux que rien du tout.

Ces pratiques recommandées ne sont pas universellement suivies (elles étaient moins couramment observées par le passé qu'elles ne le sont aujourd'hui). De plus, il n'est pas possible d'effectuer des observations parallèles lorsqu'une inhomogénéité se produit en raison d'un changement inattendu dans l'environnement de la station (par exemple, autour du site) et non dans la station elle-même. Il est possible de favoriser le respect des bonnes pratiques en intégrant la gestion des changements à l'exploitation (voir la section 1.1).

L'importance des observations parallèles pour le processus d'homogénéisation est soulignée dans la section 2.4.

### 1.3 Sauvetage des données

Le sauvetage des données est un processus continu qui consiste à recenser et préserver l'ensemble des données et métadonnées connexes, des relevés et des archives climatologiques qui risquent d'être perdus, et à numériser les données actuelles et passées sous une forme lisible par ordinateur pour pouvoir les consulter facilement. Le processus de recensement implique également de rechercher des données qui pourraient être conservées ailleurs que dans les SMHN, par exemple dans les universités, les bibliothèques et les archives nationales. Dans certains cas, les données anciennes peuvent être conservées à l'étranger. Le sauvetage des données comprend également la migration depuis des supports obsolètes ou des fichiers informatiques corrompus vers des supports modernes dans des formats lisibles.

Le sauvetage des données joue deux rôles importants dans la constitution de jeux de données climatologiques homogénéisés. Le premier rôle, le plus évident, part du fait qu'un jeu de données ne sera pas homogénéisé et analysé s'il n'est pas numérisé. De manière moins évidente, l'homogénéisation est plus efficace lorsqu'une station candidate peut être comparée à un grand nombre de stations de référence dans la région, ce qui nécessite de disposer de données numérisées pour ces stations de référence ainsi que pour la station candidate. Les stations de référence sont également utiles pour le contrôle de la qualité, tout comme les observations infraquotidiennes (par exemple, les observations horaires) dans la station candidate.

Le chapitre 2 explique l'importance du rapport signal-bruit (SNR) pour l'homogénéisation. Lorsque le SNR est compris comme la variance du signal de rupture divisée par la variance du bruit, il est important qu'il soit supérieur à un. Jusqu'à ce que ce niveau soit atteint, il conviendrait d'accorder la priorité au sauvetage de données supplémentaires pour la période et la région en question, pour autant qu'il existe d'autres données non numérisées. S'agissant du contrôle de la qualité et de l'homogénéisation des données, il est préférable de numériser les observations de groupes de stations à proximité d'une station candidate plutôt que de procéder à un échantillonnage uniforme. La numérisation de séries courtes peut également contribuer au contrôle de la qualité et à l'homogénéisation.

Les aspects pratiques détaillés du sauvetage des données n'entrent pas dans le cadre de la présente publication. Pour de plus amples informations, voir la publication OMM (2016).

### 1.4 Contrôle de la qualité

Le contrôle de la qualité a pour objectif de vérifier si une donnée transmise est bien représentative de ce qui devait être mesuré et si elle n'a pas été contaminée par des facteurs indépendants (OMM, 2018b).

Le contrôle de la qualité devrait être effectué avant l'homogénéisation car des valeurs aberrantes importantes peuvent affecter le processus d'homogénéisation. Parfois, le contrôle de la qualité est effectué de nouveau après l'homogénéisation, la qualité supérieure des données permettant de détecter des valeurs erronées plus discrètes. En outre, si les données (récentes) ont été soumises à un contrôle de la qualité en temps réel, il faut procéder à un contrôle de la qualité supplémentaire des données de séries chronologiques pour obtenir une qualité de données la plus uniforme possible sur l'ensemble de la période d'observation.

Le contrôle de la qualité peut également être une source d'inhomogénéités, en particulier s'agissant des données quotidiennes et de leur analyse pour détecter des changements au niveau de la variabilité du temps et des extrêmes météorologiques. Les méthodes utilisées pour le contrôle de la qualité des données ont souvent changé au fil du temps, ce qui peut être une source d'inhomogénéités si, par exemple, des données erronées n'ont pas été détectées par le passé, mais sont à présent signalées. Dans de nombreux cas, les données plus anciennes ont fait l'objet d'un contrôle de la qualité très limité; en particulier, les méthodes qui impliquent des études spatiales comparatives des données entre stations n'ont pu être mises en pratiques que depuis l'introduction des systèmes informatiques modernes. Plus récemment, le niveau d'intervention manuelle lors du contrôle de la qualité a été réduit dans de nombreux pays. Alors que dans un passé récent, un point de données signalé par un système automatique pouvait être soumis à un examen manuel, aujourd'hui, de plus en plus de contrôles de la qualité sont purement automatiques, ce qui entraîne parfois un risque accru de «faux positifs» lorsque des extrêmes valables sont signalés comme suspects.

Les séries chronologiques dérivées peuvent présenter des problèmes de contrôle de la qualité qui n'étaient pas manifestes dans les observations individuelles. Il est souhaitable d'effectuer un contrôle de la qualité à différentes échelles de temps afin de détecter toute la gamme des modes d'erreur possibles.

Des explications sur les tests et méthodes spécifiques au contrôle de la qualité n'entrent pas dans le cadre de cette publication. Pour obtenir davantage de détails sur le sujet, le lecteur est invité à consulter les publications OMM (1993) et OMM (1986). Les documents d'orientation étaient en cours de mise à jour lors de la rédaction de la présente publication. Pour un bon exemple de contrôle de la qualité applicable à un jeu de données mondial, voir Dunn *et al.* (2012).

## 1.5 **Historique et métadonnées des stations**

Les sous-sections suivantes décrivent le lien entre les métadonnées passées et les activités d'homogénéisation actuelles. Les normes relatives aux métadonnées et identifiants de stations modernes ont été définis par le Système mondial intégré des systèmes d'observation de l'OMM (WIGOS) (voir OMM, 2019).

### 1.5.1 **Que sont les métadonnées ?**

Dans le contexte de l'homogénéisation, le terme «métadonnées» est utilisé pour désigner ce qui, dans d'autres contextes, est appelé «métadonnées relatives à une station». L'historique d'une station est particulièrement important pour l'homogénéisation. Il comprend des informations sur le site d'observation et les instruments, ainsi que sur les procédures d'observation et de traitement des données appliquées tout au long de l'existence de la station. Ces métadonnées peuvent inclure, entre autres:

- L'emplacement et l'altitude de la station et les dates des changements qui y ont été apportés;
- Le type et l'état des instruments utilisés dans un lieu donné;
- Les dates de remplacement, par exemple, des instruments et des abris;
- L'utilisation des sols et le type de végétation à proximité de la station;
- Les changements intervenus alentour (implantation);
- Les heures et la fréquence standard auxquelles les observations sont effectuées;

- Le nom du ou des observateurs (pour les stations manuelles);
- Les procédures de traitement des données (par exemple, la définition de la température moyenne journalière);
- Les dates des étalonnages ou des contrôles de tolérance effectués à la station, et les résultats obtenus;
- Les opérations détaillées de maintenance (programmées ou non) menées à la station.

### 1.5.2 **Valeur et limites des métadonnées**

Les métadonnées, lorsqu'elles existent, sont extrêmement précieuses pour l'homogénéisation des données. Si les méthodes statistiques de détection des inhomogénéités peuvent fournir des preuves solides qu'une rupture d'homogénéité peut se produire, les métadonnées peuvent en indiquer la cause et même en préciser la date (si elle a été enregistrée) avec une grande précision. Les méthodes statistiques ne peuvent que déterminer une période, avec une précision limitée (allant généralement de quelques mois à un an). Les métadonnées peuvent également permettre de détecter des inhomogénéités dans des situations où les méthodes statistiques ont une efficacité limitée ou nulle – par exemple, lorsqu'il n'y a pas de stations de référence à proximité, ou lorsqu'un changement a des répercussions simultanées sur une grande partie des éléments du réseau.

L'homogénéisation est normalement plus efficace lorsqu'elle associe métadonnées et méthodes statistiques.

Il n'y a pas de consensus sur la question de savoir si les métadonnées devraient jouer un rôle dans la décision de confirmer ou non une rupture constatée par homogénéisation statistique. Il est d'ailleurs potentiellement dangereux de le faire car toutes les ruptures ne sont pas aussi bien documentées les unes que les autres. L'exemple le plus connu est celui d'une station urbaine où l'urbanisation progressive n'a généralement été que peu documentée, contrairement aux changements d'emplacement de la station. Ne supprimer que les changements d'emplacement qui ont été consignés amplifierait en théorie les erreurs de tendance. D'autre part, s'il y a plusieurs ruptures dans les séries à traiter et les séries de référence, l'homogénéisation peut représenter un problème combinatoire épineux. Dans de tels cas, les ruptures probables connues peuvent aider à résoudre cette énigme.

Les métadonnées sont d'autant plus précieuses qu'elles sont complètes. De nombreux Services météorologiques nationaux ont recueilli des métadonnées de bonne qualité ces dernières années, mais au fur et à mesure que l'on remonte dans le temps, les métadonnées se font souvent de plus en plus rares. En outre, les métadonnées plus anciennes sont davantage consignées sur papier, ce qui signifie qu'elles peuvent être difficiles à localiser ou à utiliser. Comme la plupart des stations ont été installées pour prendre en charge les prévisions météorologiques plutôt que les applications climatologiques, en particulier avant les années 1990, l'enregistrement de métadonnées qui n'étaient pas utiles pour les prévisions météorologiques ne constituait souvent pas une priorité élevée. En outre, les relevés de métadonnées fournissent parfois un instantané dans le temps et n'indiquent pas la date d'un changement (par exemple, deux rapports d'inspection de station à cinq ans d'intervalle peuvent indiquer que la station est à des emplacements différents sans préciser la date exacte du déménagement).

Certains aspects des métadonnées sont également mieux étayés que d'autres. L'expérience montre qu'un dossier est souvent tenu sur les nouveaux instruments alors que la documentation sur l'environnement du site, en particulier au-delà du voisinage immédiat de l'abri où sont installés les instruments, est souvent limitée ou inexistante.

Les métadonnées peuvent prendre la forme d'un point de données unique (par exemple, un ensemble de coordonnées de la station) ou d'une information plus complexe (par exemple, une photo de la station qui renseigne sur le type de surface terrestre et les obstructions environnantes). Même les formes les plus simples de métadonnées comportent un certain degré d'incertitude. Jusqu'à récemment, les coordonnées des stations étaient rarement enregistrées avec une précision suffisante pour relever les petits déplacements du site (en dizaines de mètres), alors que ces déplacements peuvent quand même être significatifs sur le plan climatologique, en particulier dans le cas d'une topographie complexe ou s'ils ont une incidence sur l'exposition du site au vent. Dans les zones côtières exposées, par exemple, on a constaté que les déplacements de moins de 50 m modifiaient de 15 à 25 % les valeurs des précipitations observées.

Dans les jeux de données plus anciens, ou ceux échangés au niveau international, les coordonnées de la station ne sont souvent communiquées qu'à la minute ou au centième de degré près, ce qui ne permet de déterminer la position de la station qu'à environ 2 km près.

Les métadonnées peuvent nécessiter une interprétation considérable, lorsque de multiples informations sont utilisées pour parvenir à une conclusion. Séparer les métadonnées qui ont un lien avec le climat de celles qui n'en ont pas peut être un processus chronophage.

Tout comme les données, les métadonnées peuvent parfois être erronées. Il est donc utile de confirmer une inhomogénéité avec plusieurs faisceaux de preuves lorsque cela est possible.

### 1.5.3 **Identifiants des stations**

Déterminer exactement à quelle station un jeu de données est associé est une partie importante de la constitution de jeux de données homogènes portant sur de longues périodes. Dans la plupart des pays, une station aura un identifiant national et les stations qui transmettent des observations sur le plan international auront également un identifiant de station de l'OMM. Il peut arriver que deux stations ou plus partagent le même identifiant de station de l'OMM. D'autres identifiants peuvent également être utilisés, tels que le nom de la station ou le code de l'Organisation de l'aviation civile internationale (OACI) pour un site aéroportuaire.

Il est bon d'associer un identifiant à une station et, si cette station subit un changement substantiel sur le plan climatologique (par exemple, un déplacement important), de cesser d'enregistrer des données sous cet identifiant et d'en créer un nouveau pour le nouvel emplacement. Cela maximise la visibilité du changement, tout en continuant de permettre la constitution d'un jeu de données à long terme avec des identifiants multiples (les identifiants multiples sont, bien sûr, nécessaires si des observations parallèles sont effectuées).

Il a été cependant courant de conserver un identifiant unique de station, même en cas de changement important. Cette pratique était particulièrement répandue pour les données anciennes, lorsque l'on avait moins conscience qu'aujourd'hui de l'impact sur le plan climatologique que les changements apportés au site pouvaient avoir.

Il est à noter que l'identifiant de station de l'OMM n'est pas considéré comme un identifiant de station unique dans certaines archives nationales. Il était initialement attribué à des fins de prévision météorologique et son nombre de chiffres est limité. Deux stations ou davantage peuvent ainsi partager le même identifiant OMM. En outre, une nouvelle station réutilise parfois l'identifiant d'une station fermée. Il existe aussi des cas où les identifiants nationaux ont été réutilisés.

Les jeux de données internationaux posent des problèmes particuliers en matière d'identification des stations. En général, ces jeux de données contiennent des données provenant de sources multiples qui peuvent être des duplications partielles les unes des autres, avec des métadonnées qui se limitent souvent aux coordonnées (qui peuvent être d'une précision limitée) et au nom de la station (qui peut avoir plusieurs orthographes, en particulier lorsqu'il contient des caractères non ASCII ou si le nom original est dans une langue qui n'utilise pas l'alphabet romain). La publication de Rennie *et al.* (2014) détaille les procédures utilisées pour fusionner les données provenant de sources multiples dans un seul grand jeu de données international: la banque de données de l'Initiative internationale concernant les températures de surface (ISTI).

### 1.5.4 **Formats et accessibilité des métadonnées**

Les métadonnées peuvent se présenter sous des formes très diverses. Dans de nombreux Services météorologiques nationaux, les métadonnées récentes sont conservées en tout ou partie sous forme numérique dans une base de données consultable, mais de grandes quantités de métadonnées anciennes n'existent généralement que sur papier.

Les métadonnées peuvent se présenter sous des formes spécifiques à une station donnée, ou dans des documents qui couvrent un grand nombre de stations (par exemple, les procédures d'observation s'appliquent souvent à tout un réseau national et sont répertoriées dans des documents nationaux plutôt que dans des documents spécifiques à une station). Dans de nombreux pays, les stations sont régulièrement inspectées par les responsables du réseau, et les



rapports d'inspection établis constituent une part importante des métadonnées disponibles. Parfois, les observateurs, leur famille et les journaux, documents comptables ou autres documents externes peuvent fournir des métadonnées supplémentaires.

Avoir accès aux métadonnées est souvent un défi considérable. Des quantités importantes de métadonnées peuvent ne pas avoir été numérisées et n'être disponibles que sur papier. Les métadonnées doivent être incluses dans les opérations de sauvetage des données, car elles sont aussi importantes que les relevés climatologiques. La plupart des métadonnées sont spécifiques à un lieu donné et peuvent être localisées dans des fichiers indexés par lieux. D'autres métadonnées pertinentes, en particulier celles qui traitent des normes et des changements à l'échelle du réseau, peuvent figurer dans des rapports annuels ou d'autres documents qui peuvent être plus difficiles à retrouver. Pour consulter des documents papier, il peut être nécessaire de se rendre en personne dans un centre d'archives.

Enfin, les métadonnées sont normalement disponibles dans la langue locale uniquement.

## 1.6 Formation

La formation du personnel qui participe aux opérations d'homogénéisation est importante pour la qualité des données homogénéisées. Elle comporte de multiples aspects: une bonne maîtrise des concepts généraux, une compréhension du contexte statistique du processus d'homogénéisation et des conseils pratiques pour choisir les méthodes d'homogénéisation les plus appropriées et manipuler le logiciel sélectionné. De plus, l'évaluation des résultats nécessite un expert qualifié.

Au moment de la rédaction de la présente publication, les possibilités de formation pertinentes étaient les suivantes:

- La série de séminaires intitulée «Homogenization and Quality Control in Climatological Databases» (Homogénéisation et contrôle de la qualité des bases de données climatologiques), qui se tient à Budapest (Hongrie) (HMS, 1996; OMM, 1999; OMSZ, 2001; OMM, 2004; OMM, 2006; OMM, 2010; OMM, 2011; OMM, 2014; OMSZ, 2017). Ces séminaires favorisent les débats relatifs aux méthodes d'homogénéisation et mettent l'accent sur les aspects théoriques, les applications pratiques et l'évaluation de celles-ci. La plupart des actes de ces séminaires sont publiés dans la série du Programme mondial des données climatologiques et de surveillance du climat (à savoir WCDMP-41, WCDMP-56, WCDMP-71, WCDMP-75 et WCDMP-78);
  - Le cours de formation annuel intitulé «La climatologie, fondation des services climatologiques», organisé par Météo-France;
  - Les ateliers de gestion des données organisés par le Réseau des Services météorologiques européens (EUMETNET);
  - La session intitulée «Climate monitoring: data rescue, management, quality and homogenization» (Surveillance du climat: sauvetage, gestion, qualité et homogénéisation des données), qui se tient lors de la réunion annuelle de la Société météorologique européenne;
  - La session annuelle intitulée «Development of climate datasets: homogenization, trends, variability and extremes, including sub-daily timescales» (Développement des jeux de données climatologiques: homogénéisation, tendances, variabilité et extrêmes, y compris les échelles de temps infraquotidiennes), qui a lieu lors de l'Assemblée générale de l'Union européenne des géosciences;
  - La session intitulée «Climate data homogenization and climate trends/variability assessment» (L'homogénéisation des données climatologiques et l'évaluation des tendances climatiques et de la variabilité du climat), organisée dans le cadre de la réunion internationale sur la climatologie statistique (IMSC).
-

## CHAPITRE 2. L'HOMOGENÉISATION DANS LA PRATIQUE

Le présent chapitre vise à décrire les problèmes généraux liés à la constitution de jeux de données homogénéisés. L'homogénéisation est généralement plus efficace lorsqu'elle est effectuée, si possible, avec des logiciels qui ont fait leur preuve. Les logiciels spécifiques à l'homogénéisation des données sont examinés au chapitre 3.

De nombreux facteurs peuvent être à l'origine d'inhomogénéités dans un relevé climatologique. Certains n'affectent que quelques éléments climatiques à un endroit précis. Certains ne concernent qu'un seul lieu, tandis que d'autres peuvent concerner tout un réseau d'observation ou des parties importantes de celui-ci. Ce dernier scénario est important car a) il peut causer un biais à grande échelle et est donc significatif sur le plan climatologique, et b) il est difficile à éliminer par homogénéisation statistique s'il se produit sur une courte période. Les inhomogénéités peuvent avoir un cycle saisonnier prononcé et/ou être dépendantes du type de temps (c'est-à-dire dépendantes du régime).

Les causes d'inhomogénéités sont notamment les suivantes:

- **Un changement de site.** Par exemple, les premières stations ont souvent été installées dans de petites villes et des villages, puis ont été déplacées vers la périphérie ou les aéroports. Les premières stations météorologiques automatiques devaient parfois être placées à proximité des bâtiments (Menne *et al.*, 2010), alors qu'avec la technologie moderne, il est devenu plus facile d'installer les stations dans des endroits vierges ou de les déplacer du milieu d'un village vers des zones périphériques (Dienst *et al.*, 2017; Dienst *et al.*, 2019).
- **Un changement dans l'environnement local.** On peut citer, comme exemples de micro-changements environnementaux, la croissance ou la coupe de la végétation autour d'un site, ou encore la construction ou la destruction d'un bâtiment à proximité. De plus, l'arrosage de l'herbe sous les instruments peut entraîner une rupture d'homogénéité. Les premières mesures de température et de précipitations étaient souvent effectuées à une hauteur de plusieurs mètres, alors qu'aujourd'hui, la norme est de 1,5 m à 2 m. Le développement de l'environnement à plus grande échelle (urbanisation) peut provoquer un réchauffement local progressif conduisant à déplacer la station, ce qui occasionnera probablement une baisse des températures enregistrées. La modification des pratiques d'irrigation de la région peut également causer un refroidissement artificiel (Cook *et al.*, 2014). En déplaçant un anémomètre du toit d'un aéroport au niveau standard de 10 m, on constate souvent une forte baisse de la vitesse du vent de surface (Wan *et al.*, 2010).
- **Un changement d'instruments.** En ce qui concerne les précipitations, les changements de la géométrie extérieure et du pare-vent entraînent une sous-estimation des valeurs et produisent donc des inhomogénéités (Leeper *et al.*, 2015). Les changements de type d'instruments, tels que les pluviomètres, ont des répercussions sur les erreurs de mesure typiques telles que la perte par mouillage et la précision de mesure (Wang *et al.*, 2017, 2010). Pour les changements de température dans l'abri (voir par exemple, Parker, 1994; Böhm *et al.*, 2010; Buisan *et al.*, 2015) et donc de rayonnement et de mouillage, la protection du thermomètre est fondamentale. Même la position du thermomètre dans l'abri peut avoir de l'importance. On a rapporté le cas d'un abri Stevenson en plastique qui laissait entrer le soleil les jours de grande chaleur. La ventilation mécanique peut réduire les erreurs liées au rayonnement, mais elle peut aussi produire des erreurs de mouillage plus importantes et perturber la couche limite stable. Le temps de réponse des thermomètres modernes et de leurs petits abris est par nature beaucoup plus court, ce qui est particulièrement important pour les températures minimales et maximales quotidiennes en raison des fluctuations turbulentes. Le verre de certains des premiers thermomètres s'est contracté au cours des premières années (Winkler, 2009). Les erreurs d'étalonnage peuvent également provoquer des inhomogénéités. Les thermomètres au mercure ne peuvent enregistrer des températures inférieures à -39 °C, c'est pourquoi les thermomètres à minimum sont désormais des thermomètres à alcool en général. Les températures record peuvent se situer en dehors de la plage (d'étalonnage) de certains capteurs météorologiques automatiques.

- **Un changement d'observateur.** L'influence de l'observateur est particulièrement notable pour les éléments qui impliquent un certain niveau de jugement, comme les nuages ou la visibilité. Dans le cas des réseaux à caractère bénévole, un changement d'observateur peut également être le signe qu'un déplacement a eu lieu, lequel pourrait ne pas être consigné ailleurs.
- **Un changement dans les procédures d'observation.** Par exemple, une modification de l'heure d'observation, qui se produit souvent simultanément dans l'ensemble du réseau. Une telle modification est importante pour les mesures à heures fixes, mais aussi pour les températures minimales et maximales quotidiennes (Vincent *et al.*, 2009; Degaetano, 2000; Vose *et al.*, 2003; Karl *et al.*, 1986). Les modifications de maintenance peuvent également avoir leur importance, par exemple le calendrier établi pour la peinture, le nettoyage et le remplacement des abris. Avec les stations météorologiques automatiques, il est possible que l'arrêt de la ventilation dû au givrage passe inaperçu et que les dommages et salissures soient détectés plus tard. Les modifications des procédures d'étalonnage peuvent en théorie avoir des répercussions sur l'ensemble d'un réseau.
- **Une modification du traitement des données.** Dans une grande partie du monde, la température moyenne journalière est calculée à partir des températures maximales et minimales quotidiennes. Cependant, en Europe, les premières mesures étaient souvent effectuées à des heures fixes, avec un calcul de la température moyenne à partir de celles-ci, parfois en association avec la température minimale ou maximale quotidienne. Avec les stations météorologiques automatiques, il existe de nombreuses définitions différentes de la moyenne journalière. Les procédures de validation et de contrôle de la qualité, ainsi que les possibilités d'effectuer un tel contrôle, ont changé au fil du temps. S'agissant des précipitations, un autre élément clé est de savoir comment les précipitations solides ont été mesurées et converties en précipitations liquides pour l'archivage (Wang *et al.*, 2017).
- **La numérisation et les erreurs dans les bases de données.** La numérisation peut produire des inhomogénéités, par exemple, lorsqu'un signe négatif (indiqué en couleur) est oublié lors de la numérisation d'un jeu de données consigné sur papier. Les données des stations ayant un nom identique ou similaire peuvent être mélangées. Dans les bases de données mondiales, les données peuvent provenir de plusieurs sources différentes dont la provenance n'est pas claire. En conséquence, des séries de données relatives à des stations ayant des noms similaires peuvent être mélangées, ou des séries légèrement différentes provenant de la même station peuvent être conservées en tant que séries distinctes. De plus, les métadonnées, telles que l'emplacement de la station et les unités, peuvent être erronées. Les erreurs généralement relevées dans les bases de données sont des valeurs 10 fois trop élevées, notamment pour les précipitations.

Le reste du présent chapitre se compose de 12 sections principalement consacrées aux sujets suivants: l'homogénéisation physique et statistique, les rôles des séries de référence, le choix ou la composition des séries de référence, les solutions possibles en l'absence de séries de référence, la validation et la mise à jour opérationnelle des résultats de l'homogénéisation, ainsi que la documentation de la procédure d'homogénéisation et du jeu de données qui en résulte.

## 2.1 Homogénéisation physique et statistique

Le fait qu'une méthode d'homogénéisation soit dite «physique» ou «statistique» dépend généralement de la manière dont les corrections sont calculées. Parfois, la principale preuve de l'inhomogénéité des données d'une station candidate est de nature statistique: un grand saut dans une série chronologique ou un comportement nettement différent par rapport aux stations voisines. Dans de tels cas, les corrections doivent être calculées statistiquement, raison pour laquelle la méthode utilisée est appelée «homogénéisation statistique».

Parfois, les raisons de ces inhomogénéités sont connues, ce qui favorise des ajustements plus précis. L'homogénéisation physique se réfère à l'utilisation d'une relation physique entre différentes variables pour estimer les ajustements à apporter. Par exemple, on a utilisé un profil logarithmique du vent, qui représente la relation entre la vitesse du vent de surface, la hauteur

de l'anémomètre et la longueur de rugosité de la surface, pour corriger les données relatives à la vitesse du vent de surface selon les changements de hauteur de l'anémomètre (Wan *et al.*, 2010). Par ailleurs, on s'est servi d'un modèle hydrostatique, qui représente la relation entre les pressions au niveau d'une station et de la mer et les températures du thermomètre sec, pour corriger les erreurs tant dans les données de pression au niveau des stations que dans celles de la pression au niveau moyen de la mer qui sont dues à des valeurs erronées d'altitude des stations (Wan *et al.*, 2007).

Autre exemple: lorsque l'heure d'observation change, les observations faites aux deux heures de relevé peuvent être utilisées pour corriger le problème engendré (Vose *et al.*, 2003; Vincent *et al.*, 2009). On parle également d'homogénéisation physique dans ce cas, bien que la méthode utilisée pour estimer l'ajustement ne soit pas strictement physique.

Lorsqu'un déplacement de station est consigné par écrit, les modifications physiques sont connues, mais la taille du saut doit être déterminée statistiquement. Cette opération est donc aussi considérée comme une homogénéisation statistique. L'utilisation de données parallèles pour estimer la taille du saut (lors d'une relocalisation) est considérée comme une opération d'homogénéisation statistique. Il convient de se fonder sur une très bonne station de référence, ce qui devrait permettre d'obtenir des résultats plus fiables.

L'homogénéisation physique peut également inclure des estimations statistiques. Par exemple, pour les changements d'heure d'observation aux États-Unis, la NOAA a élaboré une méthode de correction basée sur les observations horaires (Vose *et al.*, 2003). Afin de pouvoir corriger les données des stations où seules des observations quotidiennes sont effectuées, on a utilisé la régression linéaire multiple pour calculer les corrections du biais des heures d'observation sur une base mensuelle à partir des relevés des stations où des observations horaires sont effectuées. Les prédicteurs étaient les coordonnées des stations (fuseaux horaires, latitudes et longitudes), l'heure d'observation, l'amplitude diurne moyenne de la température et la différence moyenne de température journalière (Karl *et al.*, 1986). Des méthodes similaires ont été développées pour le Canada afin de tenir compte d'un changement de l'heure d'observation de 00:00 à 06:00 UTC apporté en 1961 (Vincent *et al.*, 2009), tandis que des données plus récentes d'une résolution d'une minute ont été utilisées pour évaluer l'impact d'un changement de l'heure d'observation de 00:00 à 09:00 heure locale intervenu en 1964 sur certains sites australiens (Trewin, 2012).

Il est préférable d'utiliser une relation physique pour estimer les ajustements lorsque les relations physiques connues entre les variables sont suffisamment fiables pour que les ajustements correspondants donnent de meilleurs résultats que l'homogénéisation statistique, comme dans les cas susmentionnés, présentés par Wan *et al.* (2010) et Wan *et al.* (2007). Cependant, les ajustements estimés sur la base d'une relation physique nécessitent des métadonnées précises (telles que les hauteurs des anémomètres et les altitudes des stations), voire d'autres données (telles que la température du thermomètre sec et la rugosité de la surface dans les exemples ci-dessus). Ces métadonnées ou données sont rarement disponibles. L'homogénéisation physique ne peut être effectuée que lorsque la cause des inhomogénéités est connue et que les métadonnées/données correspondantes sont disponibles. Ainsi, l'homogénéisation physique représente habituellement un surcroît de travail, car elle ne peut être appliquée qu'à une partie des inhomogénéités. Par conséquent, l'homogénéisation statistique devrait toujours être appliquée pour corriger les éventuelles inhomogénéités restantes.

Lorsque les méthodes physiques et statistiques sont toutes deux possibles, il est recommandé de calculer les deux ajustements et de les comparer. Une telle comparaison peut aider à repérer les problèmes. Le choix entre l'homogénéisation physique ou l'homogénéisation statistique dépend de la précision des corrections. Dans la plupart des cas, les ajustements physiques sont plus précis, mais pas toujours. Appliquer d'abord des ajustements physiques risque de rendre les inhomogénéités restantes trop petites pour être détectées par homogénéisation statistique, mais suffisamment importantes pour poser des problèmes à grande échelle au plan climatologique. En conclusion, utiliser la méthode la plus précise peut donner de meilleurs résultats qu'une utilisation consécutive des deux méthodes.

## 2.2 **Sélection des données à homogénéiser**

### 2.2.1 **Quelles stations convient-il de sélectionner?**

Les orientations données dans la présente sous-section s'appliquent à l'homogénéisation de jeux de données et non d'une série chronologique unique.

Deux grandes approches peuvent être utilisées pour sélectionner les données à homogénéiser dans un jeu de données national ou régional.

La première approche consiste à inclure toutes les stations qui répondent aux critères prédéfinis pour des facteurs tels que la durée des relevés et l'exhaustivité des données, tandis que la seconde consiste à ne prendre en compte que les stations retenues pour la qualité de leurs observations (par exemple, les stations ayant les meilleures normes d'emplacement ou le moins de déplacements enregistrés) ou leur représentativité géographique.

Ces deux approches ont été utilisées pour les principaux jeux de données nationaux; par exemple, les jeux de données nationaux homogénéisés des États-Unis, du Canada et de l'Espagne (Menne *et al.*, 2009; Vincent *et al.*, 2012; Guijarro, 2013) comprennent toutes les stations ayant des relevés suffisamment longs, tandis que pour les jeux de données de la Suisse et de l'Australie et un autre jeu espagnol (Begert *et al.*, 2005; Brunet *et al.*, 2006; Trewin, 2013), on a choisi, dans des réseaux nationaux beaucoup plus grands, les stations dont la qualité et la représentativité étaient estimées comme les meilleures.

Une sélection rigoureuse réduit la quantité de travail, par conséquent une plus grande attention peut être accordée à chaque station. L'objectif devrait être d'obtenir un SNR, défini comme la racine carrée de la variance du signal de rupture divisée par la variance du signal de bruit (Lindau et Venema, 2018a)), supérieur à un. L'utilisation de plusieurs stations a l'avantage de permettre l'étude des climats régionaux et la production de cartes.

Des méthodes de calcul du SNR d'une série chronologique de différences sont présentées par Lindau et Venema (2018a, 2019).

### 2.2.2 **Comment préparer les données pour l'homogénéisation**

Une fois qu'un ensemble de stations a été sélectionné pour l'homogénéisation, l'étape suivante consiste à déterminer quelles données de ces stations seront utilisées dans le processus de détection des points d'inflexion. Les décisions prises ici sont importantes parce que certaines inhomogénéités peuvent avoir un cycle saisonnier – par exemple, un changement de la protection contre le vent d'un pluviomètre sur un site à climat froid peut avoir peu d'impact sur les précipitations estivales, mais un impact conséquent sur les précipitations hivernales, dont la plupart tombent sous forme de neige.

Les trois choix possibles sont les suivants:

- a) Utiliser uniquement des données annuelles (somme annuelle ou moyenne annuelle, parfois associée à l'amplitude du cycle saisonnier);
- b) Utiliser des séries chronologiques mensuelles ou saisonnières en parallèle les unes des autres (et/ou avec des données annuelles);
- c) Utiliser des données mensuelles ou saisonnières (sous leur forme originale ou sous forme d'anomalies) comme une série chronologique unique (également appelée série chronologique mensuelle ou saisonnière de données consécutives).

L'utilisation des données annuelles présente généralement le SNR le plus favorable pour les inhomogénéités qui concernent la totalité ou une partie substantielle de l'année. Par rapport aux données mensuelles ou journalières consécutives, le SNR plus faible est cependant totalement compensé par le plus grand nombre de valeurs (dans le cas du bruit blanc et pour autant que les ruptures soient au 1<sup>er</sup> janvier; Lindau et Venema, 2018a). En revanche, l'utilisation de données annuelles peut conduire à ne pas détecter les inhomogénéités qui n'affectent qu'une partie de l'année, ou les inhomogénéités qui ont des impacts opposés à des saisons différentes et qui s'annulent mutuellement dans une moyenne annuelle (par exemple, lorsqu'un site est déplacé d'un endroit côtier exposé à un autre plus à l'intérieur des terres, il est probable qu'on relève des températures plus élevées en été mais plus basses en hiver). Dans ces cas, le recours au cycle saisonnier ou à l'option b) ci-dessus peut améliorer les résultats.

Utiliser en parallèle des séries chronologiques mensuelles ou saisonnières (c'est-à-dire réaliser des tests séparément sur les séries chronologiques de chaque mois/saison avec un point de données par an) permet de détecter des inhomogénéités variant selon les saisons. Dans certains cas, une station qui ne présente pas d'inhomogénéité significative sur une échelle de temps annuelle peut montrer des signaux saisonniers significatifs, par exemple lorsque des signaux opposés en été et en hiver s'annulent dans la moyenne annuelle. Cette méthode nécessite de regrouper les informations provenant des différents tests mensuels/saisonniers et les valeurs annuelles, afin d'élaborer un ensemble final d'inhomogénéités potentielles. Cependant, prises isolément, ces séries mensuelles (ou saisonnières) parallèles sont de même longueur que les séries de données annuelles correspondantes, mais elles sont plus bruyantes. Les séries mensuelles (ou saisonnières) consécutives forment des séries chronologiques plus longues, ce qui compense en partie le bruit. Il n'est utile de tester les données mensuelles parallèles que dans le cas de réseaux denses.

Certaines méthodes, au titre desquelles des données mensuelles ou saisonnières consécutives sont considérées comme une série chronologique unique, permettent d'augmenter la résolution temporelle pour détecter les inhomogénéités et, comme les méthodes utilisant des séries mensuelles en parallèles, évitent d'affecter la totalité de l'année qui contient la rupture à la sous-période homogène avant ou après la rupture. Cependant, ces méthodes peuvent également ne pas détecter un signal variable selon la saison. Elles peuvent également être compliquées par le fait que, sous certains climats, non seulement la valeur moyenne d'une série chronologique varie selon les saisons, mais sa variabilité peut également avoir un cycle saisonnier (par exemple, dans la plupart des endroits de l'hémisphère Nord situés à des latitudes moyennes et élevées, la variabilité des températures est nettement plus importante en hiver qu'en été). En outre, l'autocorrélation dans les séries mensuelles ou saisonnières consécutives est généralement plus élevée que dans les séries annuelles. Elle n'est donc pas négligeable et devrait être prise en compte lors du test statistique.

Les méthodes utilisant des données annuelles ne permettent pas de déterminer le moment où se produit une inhomogénéité, sans référence aux métadonnées, avec une précision supérieure à un an. En revanche, si le SNR des données mensuelles ou saisonnières n'est pas assez important, la date de changement aura une incertitude évidente (Lindau et Venema, 2016). Ainsi, la plus grande précision apparente qui semble être obtenue à partir d'une méthode utilisant des séries mensuelles peut être en partie illusoire. Il est à noter qu'une plus grande précision temporelle peut aussi être utile parfois pour cibler la recherche de métadonnées.

Les méthodes les plus utilisées sont l'option a) ci-dessus, pour l'homogénéisation automatique, et l'option b) avec des données annuelles, pour l'homogénéisation manuelle. L'option b) est susceptible de détecter un éventail plus large d'inhomogénéités si elle est correctement appliquée. Il est également possible d'associer deux options ou les trois (Xu *et al.*, 2017, 2013).

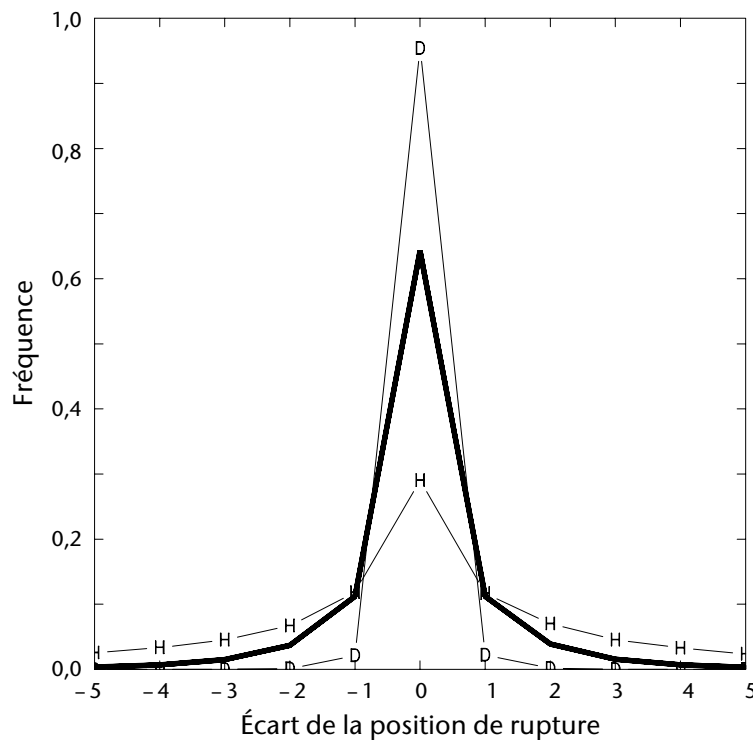
### 2.3 Détection statistique ou détermination des points de changement

Il conviendrait de toujours supposer que les métadonnées sont incomplètes et de recourir à la détermination statistique des points de rupture. Pour un test d'homogénéité, l'hypothèse nulle signifie que la série à traiter est homogène et l'hypothèse alternative, que la série à traiter n'est pas homogène (et a un ou plusieurs points de changement). En général, un test statistique consiste à comparer la valeur d'une statistique de test à sa valeur critique correspondant au niveau de signification choisi  $\alpha$  (généralement 5 %). L'hypothèse nulle est rejetée lorsque la valeur de la statistique de test dépasse la valeur critique. Le niveau de signification  $\alpha$  est la probabilité que l'hypothèse nulle soit rejetée à tort.

Même pour un point d'inflexion fixé par écrit (par exemple, lorsque l'heure du changement est consignée dans des métadonnées et connue de la personne qui analyse les données), il est nécessaire d'effectuer un test statistique pour déterminer si le changement non climatique enregistré est statistiquement différent de zéro à un niveau de signification choisi. En outre, même lorsqu'une inhomogénéité est significative sur le plan statistique, il peut subsister une grande incertitude quant à l'ampleur de l'ajustement à apporter.

Les inhomogénéités sont généralement soudaines mais elles peuvent aussi être progressives (par exemple, dans les cas de croissance de la végétation ou d'urbanisation). L'homogénéisation statistique consiste habituellement à modéliser les inhomogénéités par une fonction en escalier. Cette approche fonctionne bien pour les inhomogénéités progressives également, car elles comportent souvent des sauts supplémentaires, et comme leur comportement dans le temps peut ne pas être linéaire, l'ajustement linéaire est souvent inexact. Ainsi, dans la pratique, l'homogénéisation des inhomogénéités progressives avec de multiples ruptures fonctionne bien (Venema *et al.*, 2012).

La plupart des tests d'homogénéité sont développés pour détecter les décalages moyens et ne peuvent donc pas détecter un décalage de variance ou un changement de distribution des probabilités qui ne s'accompagne pas d'un



**Figure 2. Distribution de l'écart de la position de rupture pour trois rapports signal-bruit (SNR) différents. La ligne épaisse correspond à un SNR de 1, soit la valeur pour laquelle la variance de rupture est aussi grande que la variance de bruit. La ligne D représente le double SNR et la ligne H, la moitié du SNR (figure extraite de Lindau et Venema (2016)).**

décalage moyen. Une variante du test de Kolmogorov-Smirnov (K-S) développée par Dai *et al.* (2011) peut être utilisée pour détecter des changements inconnus dans la distribution des probabilités (y compris le décalage de variance) des données. Malheureusement, cette méthode n'a été incluse dans aucun logiciel d'homogénéisation des données. Szentimrey (2018) travaille sur une méthode visant à détecter et corriger les ruptures de l'écart moyen et de l'écart type pour des données normalement distribuées.

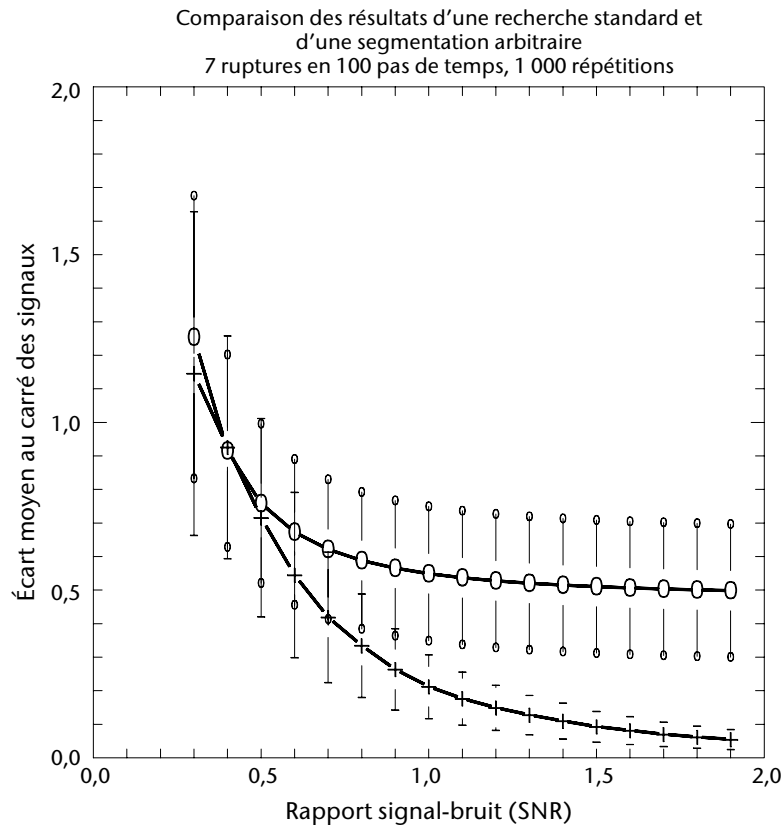
Après avoir établi une liste de points de changement qui consiste en des points de changement non climatiques significatifs, il faut estimer les ajustements à apporter pour homogénéiser la série candidate. Cette question est approfondie dans la section 2.6. La détection et la correction sont principalement effectuées par comparaison avec les stations voisines. Pour plus d'informations sur les séries de référence, voir la section 2.4. Le lecteur est invité à se reporter au chapitre 3 afin d'obtenir une description détaillée des progiciels d'homogénéisation des données climatiques qui visent à détecter et tester les points de changement et à estimer les ajustements à apporter.

### 2.3.1 **Intégration des métadonnées dans les tests statistiques**

Un test de rupture statistique fournit une estimation de la position du point d'inflexion. Sa précision dépend du SNR. Si le SNR est bien supérieur à un, la position de rupture peut être estimée avec précision (Lindau et Venema, 2016).

Se documenter sur une rupture peut prendre beaucoup de temps. Les preuves statistiques d'une rupture peuvent réduire la longueur de la période pour laquelle des métadonnées doivent être étudiées. Le SNR indique l'étendue que cette période devrait avoir. Pour les petits SNR, des erreurs de plusieurs points de données avant ou après le point d'inflexion réel sont possibles, comme le montre la figure 2.

Lorsque le SNR est très faible, une segmentation aléatoire de la série chronologique peut même éclairer autant sur le signal de rupture réel que la segmentation estimée par une méthode



**Figure 3. Précision de l'estimation du signal de rupture en fonction du rapport signal-bruit (SNR). La courbe accompagnée de cercles montre les résultats d'une méthode de détection des points de rupture tandis que l'autre courbe montre les résultats de segmentations aléatoires. Lorsque le SNR est élevé (partie droite du graphique), le signal de rupture est estimé avec précision par les méthodes de détection de rupture (signes «+»): l'écart mis au carré entre le signal détecté et le signal inséré est faible. Une segmentation aléatoire peut cependant expliquer la moitié de la variance du signal de rupture. Lorsque le SNR est inférieur ou égal à 0,5, la segmentation de la méthode d'homogénéisation n'est pas meilleure que la segmentation aléatoire. Comme les ruptures détectées expliquent à la fois la variance de bruit et la variance de rupture, la variance est plus importante que prévu dans le cas du bruit, et la détection est statistiquement significative. Le problème est que l'algorithme détecte correctement la présence d'inhomogénéités dans les observations mais ne parvient pas à déterminer leurs positions (voir Lindau et Venema (2018a) pour plus de détails).**

d'homogénéisation (voir la figure 3). En d'autres termes, le test détecte à juste titre que la série présente des inhomogénéités, mais il est possible que les positions estimées de celles-ci soient déterminées davantage par le bruit que par le signal de rupture.

Lorsque l'on associe des points de rupture estimés statistiquement avec des ruptures documentées, il faut garder à l'esprit que les métadonnées peuvent être erronées. La probabilité que les métadonnées soient erronées et l'incertitude des données dépendent de la situation et sont subjectives. En général, si les métadonnées contiennent une date spécifique, elles sont plus précises, mais si les preuves statistiques sont solides, elles peuvent prévaloir. Si une rupture détectée statistiquement se situe à proximité d'un changement consigné sans date précise (c'est-à-dire dans l'incertitude de la détection de la rupture), on peut attribuer la rupture à ce dernier.

Les tests statistiques permettant de détecter les points de changement inconnus sont différents de ceux permettant de déterminer la signification statistique des points de changement documentés. En cas de point de rupture inconnu, chaque position est testée (problème de tests multiples) et la différence maximale attendue sous l'hypothèse nulle est donc plus importante. Les tests couramment utilisés pour détecter un ou plusieurs points de changement inconnus (décalages moyens) sont des tests *t* maximaux – par exemple, le *Standard Normal Homogeneity Test* d'Alexandersson (1986) et le *Penalized Maximal t Test* de Wang *et al.* (2007) – ou des tests *F* maximaux, tels que le *Penalized Maximal*



*F Test*) de Wang (2008b), et les tests de Lund et Reeves (2002) reposant sur le modèle de régression à deux phases. Les méthodes modernes de détection des points de rupture multiples, telles que PRODIGE, MASH, ACMANT et HOMER, testent efficacement toutes les combinaisons de rupture et prennent donc également en compte les tests multiples (voir le chapitre 3 pour plus de détails sur les méthodes mentionnées dans le présent paragraphe).

Les tests couramment utilisés pour déterminer la signification des points de changement enregistrés sont le test *t* standard de Student, avec une série de référence, et le test *F* standard, lorsque l'on teste une série sans recourir à une série de référence ou lorsque l'on suspecte une différence de tendance entre la série à traiter et la série de référence (Lund et Reeves, 2002; Wang, 2003; Wang, 2008a). Comme expliqué ci-dessus, les valeurs critiques pour le test *t* ou *F* standard sont bien inférieures à celles du test *t* ou *F* maximal correspondant (voir, par exemple, Lund et Reeves (2002) et Wang (2003)). Ainsi, l'utilisation du test *t* ou *F* standard pour détecter des points de changement inconnus entraînerait trop de fausses alertes (c'est-à-dire de nombreux signalements de points de changement qui sont en fait insignifiants).

Lorsque des métadonnées sont disponibles, on peut d'abord effectuer un test statistique pour détecter les points de changement significatifs inconnus; on ajoute ensuite tous les points de changement supplémentaires enregistrés pour tester leur signification statistique. Cette étape est nécessaire parce que certains changements apportés au site, à la procédure ou aux instruments d'observation et documentés peuvent ne pas induire de changement significatif dans les séries de données candidates, alors que les points de changement identifiés statistiquement devraient être pris en compte lors du test des points de rupture consignés qui prend en compte la sous-période homogène avant et après la rupture notée.

Le test des points de rupture fixés par écrit ne devrait être utilisé que pour les points de rupture ayant un nombre limité de causes clairement justifiées et très probables, telles que les déplacements et les changements de type d'abri. Il ne devrait pas être utilisé pour des événements qui se produisent régulièrement, comme la maintenance ou l'étalonnage. Il est à noter que le fait d'inclure trop de ruptures documentées dans les tests réduira la taille de l'échantillon (c'est-à-dire que les segments seront plus courts) et augmentera donc l'incertitude des résultats. Des tests sont aussi couramment effectués pour les points de rupture inconnus, ce qui équivaut à augmenter considérablement le niveau de signification (Lund et Reeves, 2002; Wang, 2003).

## 2.4 Séries de référence

Il est plus difficile de détecter et de corriger une inhomogénéité dans une série chronologique de données climatologiques car ce type de série contient un bruit important, qui résulte de la variabilité naturelle du climat et du temps ainsi que d'erreurs de mesure. Les inhomogénéités peuvent donc être difficiles à repérer. Par exemple, il sera difficile, voire impossible, de détecter avec une méthode statistique une inhomogénéité de 0,4 °C dans une série chronologique avec un écart type de 1 °C.

Une méthode courante pour résoudre ce problème consiste à comparer la série chronologique de la station candidate à celle d'une station de référence (voisine). L'exemple le plus connu est le calcul d'une série chronologique de différences entre la série à traiter et la série de référence, puis la détection des inhomogénéités et l'ajustement de la série en fonction de cette série de différences. Les étapes à suivre dans le cas où aucune station de référence n'est disponible sont examinées au point 2.5 ci-dessous.

La variabilité naturelle qui existe dans une série candidate se retrouvera également dans la série de référence. Par conséquent, en créant une série qui compare la série candidate et la série de référence, on ôte une grande partie de l'influence de la variabilité climatique naturelle tout en conservant l'effet de l'inhomogénéité sur le site candidat.

Autre avantage des séries de référence, aucune hypothèse sur la nature statistique de la variabilité climatique n'est nécessaire, puisque l'on utilise une série de référence représentative. On peut supposer que les séries chronologiques de différences contiennent du bruit blanc (ou bruit auto-corrélé) et des inhomogénéités, ce qui simplifie grandement le problème statistique. En règle générale, une série de référence comprendra des données provenant d'un ou plusieurs endroits proches de la station candidate.

Le meilleur ensemble possible de série candidate et de série de référence est une observation parallèle de l'ancienne et de la nouvelle situation sur le même site, par exemple lorsqu'un nouveau système d'instruments est mis en place alors que l'ancien système continue de fonctionner pendant un certain temps.

Les observations parallèles sont d'autant plus utiles que la «vienne» partie du système d'observation parallèle est représentative des conditions qui prévalaient avant le début de la période d'observation parallèle. Cependant, il est courant que l'ancien environnement du site ait été modifié pendant la période d'observations parallèles, ce qui rend les observations parallèles non représentatives de cet ancien environnement. La comparaison avec les données des stations voisines pourrait aider à déceler ce problème et est donc fortement recommandée.

S'il n'y a pas de mesures parallèles disponibles, on peut envisager de les réaliser avec l'équipement utilisé avant et après la rupture ou, si celui-ci n'est pas disponible, avec des répliques créées pour l'expérience. Cette approche est particulièrement recommandée pour étudier l'influence d'une transition historique qui a eu des répercussions sur une grande partie, ou sur la totalité, du réseau (Brunet *et al.*, 2011; Mekis et Vincent, 2011; Quayle *et al.*, 1991).

Une série de données de référence peut être un composite calculé à partir de plusieurs stations voisines (référence composite) ou d'une seule station voisine (homogénéisation par paire). Dans ce dernier cas, les inhomogénéités de la série de comparaison peuvent appartenir à la série à traiter comme à la série de référence et de multiples paires doivent être étudiées pour déterminer à quelle station appartient la rupture.

Il y a quatre éléments à prendre en compte pour la sélection ou la pondération des séries de référence:

- a) Le jeu de références doit couvrir toute la période de la série à traiter;
- b) La pondération devrait réduire le bruit des séries de différences;
- c) L'influence des inhomogénéités de la référence devrait être réduite;
- d) La similarité du signal climatique régional de la série à traiter et de la série de référence devrait être renforcée.

Ces quatre éléments s'opposent les uns aux autres et la solution optimale n'est pas évidente. Par conséquent, de nombreuses méthodes différentes sont utilisées pour sélectionner les stations de référence et pour pondérer les stations voisines lors du calcul d'une référence composite à partir de celles-ci. Les méthodes de pondération courantes utilisent des facteurs de pondération basés sur des corrélations, les corrélations de la première série de différences, le krigeage (interpolation optimale), l'inverse de la distance de la station candidate et la différence d'altitude. De plus, le fait de donner le même poids à toutes les stations de référence permet d'éviter l'influence excessive d'une station voisine très inhomogène mais très proche ou très bien corrélée.

Les stations peuvent être exclues lorsque la distance ou la différence d'altitude est trop importante ou la corrélation trop faible pour sélectionner des références ayant un climat similaire.

Dans de nombreux cas, l'utilisation de la distance est similaire à l'utilisation de corrélations et la corrélation entre deux stations va diminuer de façon raisonnablement monotone avec la distance qui les sépare, mais cela peut, par exemple, ne pas être le cas en montagne, où l'altitude est également importante. La matrice de corrélation peut également être anisotrope; par exemple, les températures d'une station côtière exposée sont susceptibles d'être mieux corrélées avec une autre station exposée de façon similaire 100 km plus loin sur la côte qu'avec une station située 50 km à l'intérieur des terres.

Pour ce qui est de la corrélation entre deux stations, il faut indiquer la série à traiter et la série de référence potentielle par  $B(t)$  et  $R(t)$ , respectivement. Les séries  $\Delta_B(t) = [B(t) - B(t-1)]$  et  $\Delta_R(t) = [R(t) - R(t-1)]$  sont appelées «séries de différences premières». La corrélation entre les séries de différences premières  $\Delta_B(t)$  et  $\Delta_R(t)$  est souvent utilisée pour sélectionner une série de référence, car cette valeur de corrélation sera beaucoup moins affectée par les inhomogénéités qui pourraient exister dans la série à traiter et/ou la série de référence (une inhomogénéité générera une seule valeur incorrecte dans la série de différences premières, contre un segment de valeurs incorrectes dans la série à traiter ou de référence).

### 2.4.1 **Chevauchement**

Les données de périodes antérieures sont généralement plus rares et il est difficile de trouver des stations de référence bien corrélées. Afin de disposer de stations de référence pour des périodes antérieures, il se peut qu'il faille inclure, dans l'ensemble des stations de référence, des références présentant des corrélations plus faibles.

Le nombre minimum théorique de stations nécessaires pour procéder à une homogénéisation statistique est de trois. En pratique, cinq stations (soit quatre stations de référence) sont nécessaires pour obtenir de bons résultats dans des situations plus complexes. Cette exigence conditionne généralement l'année de début d'un jeu de données homogénéisé.

Généralement, les réseaux disposent de moins de données en début (et parfois en fin) d'exploitation. Le fait d'avoir des stations qui se chevauchent pour la première période signifie normalement qu'il faudra choisir des stations avec des corrélations plus faibles. Lorsque l'on ajoute des séries plus courtes à une référence composite, le début ou la fin de la série chronologique peut introduire une inhomogénéité. Des problèmes similaires peuvent survenir lorsqu'il manque des données pour certaines périodes.

Lors de la sélection ou de la pondération des références en fonction de leurs corrélations, il conviendrait de tenir compte du fait que les corrélations calculées comportent des incertitudes considérables, en particulier dans le cas de séries courtes. Par conséquent, la forte corrélation d'une série courte peut être une coïncidence et une série plus longue avec une corrélation plus faible peut être plus fiable.

Les séries courtes qui ne portent que sur quelques années sont également problématiques, car la détectabilité d'une rupture dépend de la capacité de détecter une différence statistiquement significative de la valeur moyenne avant et après une rupture. Ainsi, outre la taille de l'inhomogénéité, le nombre d'échantillons, c'est-à-dire la longueur des sous-périodes homogènes, est également important.

### 2.4.2 **Réduction du bruit**

Le krigeage fournit une estimation optimale des observations à la station candidate, compte tenu des observations des références. Une référence composite calculée comme une moyenne pondérée des références à l'aide de poids de krigeage réduira de manière optimale le bruit des séries chronologiques de différences. Les poids de krigeage sont calculés à l'aide de la matrice de corrélation croisée.

En raison des trois autres considérations mentionnées ci-dessus et pour simplifier leur utilisation, les corrélations elles-mêmes, ou l'inverse de la distance, sont également souvent utilisées pour calculer les poids d'une série de référence composite.

La détectabilité d'une inhomogénéité est principalement fonction du SNR. Dans le cas d'un point de rupture, le SNR est souvent défini comme le rapport entre la taille de cette inhomogénéité et l'écart type de la série chronologique considérée. Réduire la quantité de bruit dans une série chronologique en utilisant une référence augmentera la probabilité de détecter une inhomogénéité d'une taille donnée.

Comme le montre la figure 4, dans le cas d'un seul point d'inflexion dans une série chronologique de données de longueur  $N = 600$ , la probabilité de détecter de manière raisonnablement précise les inhomogénéités passe d'environ 53 % lorsque le rapport  $\Delta/SD = 0,5$  à environ 91 % (99 %) lorsque  $\Delta/SD = 1,0$  (1,5); elle est  $\geq 99,99$  % quand  $\Delta/SD \geq 2,0$ . Un SNR plus faible est également associé à une plus grande plage d'incertitude de la puissance de détection. Voir la section 2.3.1 pour une argumentation similaire concernant le cas des points de rupture multiples. Il est donc hautement prioritaire d'essayer d'obtenir un SNR supérieur à 1. Pour des valeurs SNR plus élevées, les trois autres considérations susmentionnées (voir la section 2.4) deviennent importantes.

Deux stations en homogénéisation par paires présenteront deux fois plus de ruptures qu'une série chronologique de différences basée sur une référence composite, car cette dernière offre une meilleure estimation du signal climatique régional qu'une station unique. Ainsi, si le SNR est faible, l'utilisation de références composites peut être préférable.

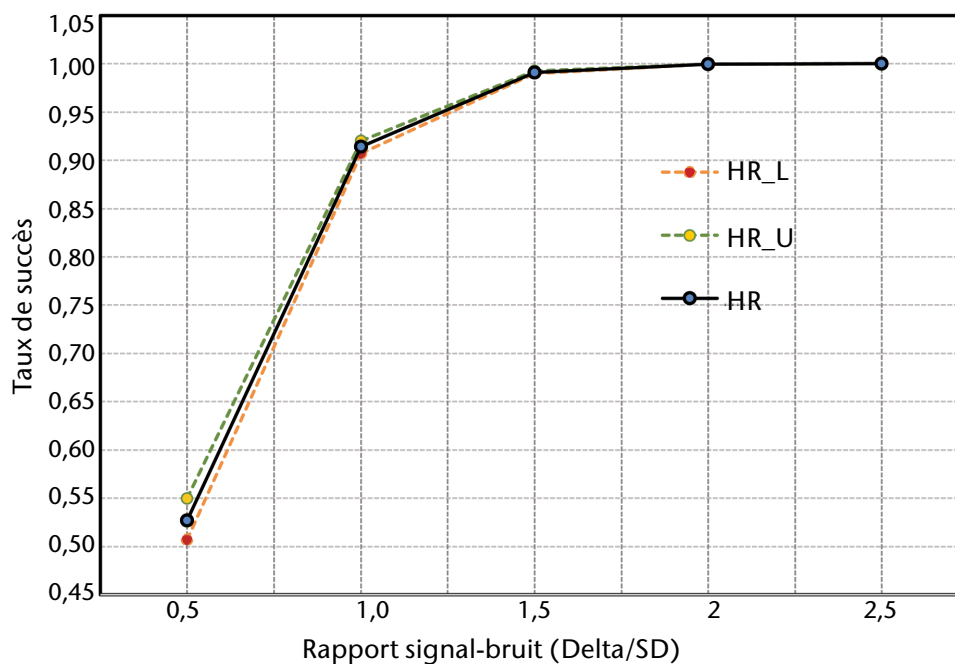


Figure 4. Puissance de détection, c'est-à-dire taux de succès (HR, axe des ordonnées) en fonction du rapport  $\Delta/SD$  (axe des abscisses) de la taille  $\Delta$  d'une inhomogénéité par rapport à l'écart type (SD) de la série candidate (sur la base du tableau 1e de Wang (2008a)). Les limites inférieure (HR L) et supérieure (HR U) représentent l'intervalle de confiance de 95 % (plage d'incertitude) des taux de succès.

#### 2.4.3 **Référence contenant des inhomogénéités**

Une série de référence devrait être homogène, en tout cas autour de la période où la série à traiter ne l'est pas. Les inhomogénéités de la série de référence peuvent facilement être prises pour des inhomogénéités de la série à traiter. On peut considérablement réduire de telles erreurs en visualisant les séries de différences ou les ajustements de régression. Pour réduire le risque qu'une référence composite présente des inhomogénéités significatives sur la période considérée, un nombre suffisant de stations de référence doit avoir des poids significatifs.

En cas d'inhomogénéités généralisées se produisant pour de nombreuses stations ou pour l'ensemble de celles-ci sur une courte période, il faut être très prudent lors du calcul de la référence composite. Cette dernière peut ne pas contenir de sauts visibles, mais si l'inhomogénéité généralisée a un biais, celui-ci se retrouvera dans la référence composite. Cette situation est particulièrement problématique pour le calcul des corrections.

Il est possible d'atténuer ce problème en recourant à des procédures itératives soigneusement élaborées, qui écartent les stations de référence présentant des ruptures lors du calcul de la référence composite ou qui corrigent les ruptures dans les stations de référence. Les méthodes d'homogénéisation par paires sont les mieux adaptées pour traiter ces cas difficiles, mais d'autres méthodes qui passent par des processus itératifs de correction des ruptures donnent également de bons résultats.

#### 2.4.4 **Références avec des signaux climatiques similaires**

Les personnes les mieux placées pour sélectionner des références avec des signaux climatiques similaires sont les experts qui connaissent le climat local. S'agissant des méthodes automatiques, la similarité est souvent estimée par des corrélations croisées ainsi que par des seuils de distance et de différence d'altitude maximales et de corrélation minimale.

Cependant, une corrélation élevée ne garantit pas que les stations de référence proviennent de la même région climatique même si elles partagent un signal similaire. D'autres critères objectifs à prendre compte pourraient être les climats de Köppen, la durée du cycle saisonnier, le cycle journalier, l'exposition, l'humidité du sol et la végétation.

Dans le cas de réseaux à faible densité, il peut être nécessaire de limiter le nombre de stations de référence afin de s'assurer qu'elles appartiennent toutes à la même région climatique.

Avec les méthodes de détection qui consistent à compter le nombre de ruptures détectées, il est particulièrement important de vérifier que toutes les stations ont un signal climatique similaire. Une telle étape est parfois utilisée pour rassembler les résultats de plusieurs tests de détection de ruptures ou lors de l'attribution de méthodes d'homogénéisation par paires, lorsque la station présentant la rupture est repérée en fonction du nombre de paires qui présentent une rupture. Si toutes les paires n'appartiennent pas à la même région climatique, il est possible que de nombreuses stations de référence éloignées laissent penser, à tort, qu'il y a une rupture, alors que cela ne se produirait pas avec un nombre plus restreint de stations de la même région climatique. Dans ce cas, il convient de sélectionner un nombre inférieur de stations de référence ou d'appliquer des pondérations plus élevées aux stations voisines.

#### 2.4.5 **Utilisation des références**

Les références sont principalement utilisées comme suit: a) la différence entre la candidate et la référence est utilisée pour les variables normalement distribuées, et b) le rapport entre la candidate et la référence est utilisé pour les variables approximativement distribuées log-normalement.

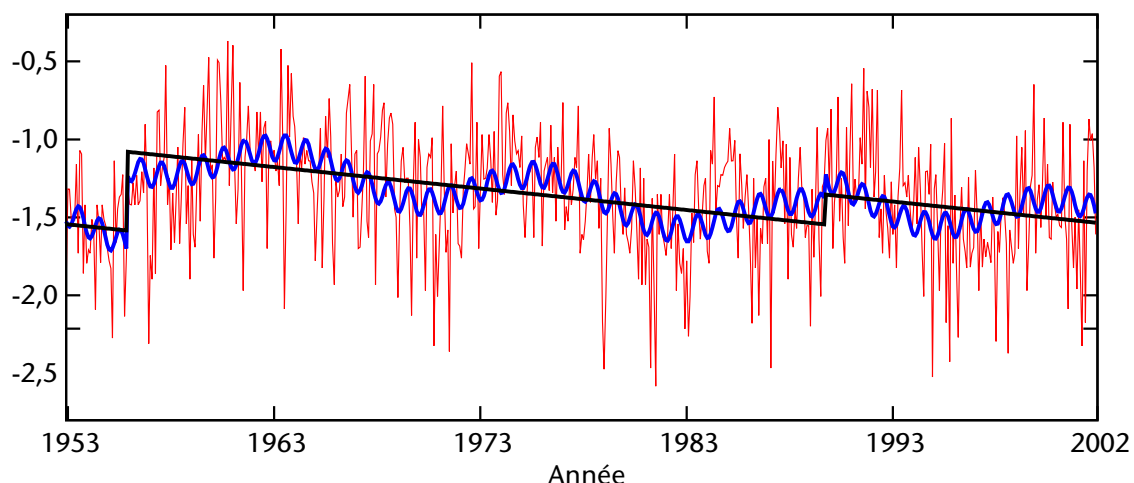
Dans le cas des séries de différences, le test est appliqué à la série de différences  $D(t) = [X(t) - Y(t)]$  où  $X$  désigne la station candidate et  $Y$  désigne la série chronologique (ou station) de référence, et les ajustements sont estimés à partir de la série de différences. La série de rapports serait calculée comme suit:  $R(t) = [X(t) / Y(t)]$ . La température et la plupart des autres variables proviennent de processus additifs et ont une distribution approximativement normale. Les précipitations mensuelles et la vitesse du vent sont des processus multiplicatifs, qui produisent une distribution approximativement log-normale. Sous certains climats, même les moyennes mensuelles de ces variables peuvent contenir des zéros. Dans ce cas, on peut transformer les données en une distribution approximativement normale et utiliser une série chronologique de différences.

L'approche covariable est moins courante. Elle se réfère à l'utilisation de la série de référence comme covariable dans un test basé sur la régression, tel que le test de Vincent (1998) basé sur la régression multiple. Dans ce cas, le test est appliqué à la série résiduelle de la régression de la série candidate par rapport à la série de référence, par exemple, les résidus  $\varepsilon(t) = [B(t) - \hat{B}(t)]$  de la régression ajustée  $\hat{B}(t) = [\hat{a} + \hat{b}R(t)]$ .

#### 2.5 **Solutions possibles en l'absence de stations de référence utiles**

Pour certaines séries d'observation, comme celles du nord du Canada (où la plupart des sites sont distants de 400 à 800 km les unes des autres), d'îles éloignées ou de l'Antarctique, il n'y a pas de référence utile disponible. C'est le SNR qui importe pour l'homogénéisation et non la distance. Pour ce qui concerne le vent, les précipitations ou l'humidité, le SNR sera faible à des distances beaucoup plus courtes que pour la température et, surtout, la pression au niveau de la mer. Ces stations éloignées sont importantes et ont un grand poids dans les séries de données moyennes régionales ou mondiales, car elles représentent une grande surface et fournissent souvent des données sur des périodes anciennes qui présentent un intérêt scientifique. Ces situations nécessitent l'avis d'experts et il est préférable de les traiter en récupérant et en étudiant les métadonnées. Les approches ci-après peuvent être suivies pour homogénéiser les données provenant de ces sites éloignés.

Premièrement, on peut chercher d'autres variables ou sources de données connexes à utiliser comme référence. Par exemple, pour les stations côtières et insulaires, les températures de surface de la mer peuvent constituer une autre série de référence possible (Cowtan *et al.*, 2018). Cependant, les températures de l'air et de la surface de la mer ont des tendances et une variabilité différentes. La couverture nuageuse a été utilisée pour homogénéiser les observations sur la durée d'insolation, et les données sur la durée d'insolation ont servi à homogénéiser



**Figure 5. La série chronologique de Fort Nelson (Canada) sur la nébulosité (rouge) et le signal et les écarts climatiques (courbe bleue) répertoriés grâce à la méthode de Wen *et al.* (2011). La ligne noire épaisse montre la tendance et les écarts, à l'exclusion des composantes périodiques (extrait de la figure 8 de Wen *et al.* (2011)).**

les données sur le rayonnement en surface (Yang *et al.*, 2018). Il faut veiller à ce que cela ne supprime pas la variabilité de la durée d'insolation due aux variations de la microphysique des nuages et des aérosols. Les températures de l'air en surface provenant d'un jeu de données de réanalyse peuvent être utilisées comme référence pour tester la température de l'air en surface relevée à un site distant, en gardant à l'esprit que les données de réanalyse présentent des inhomogénéités qui leur sont propres.

Deuxièmement, il existe des méthodes d'homogénéisation qui peuvent être utilisées sans recourir à une série de référence (homogénéisation statistique absolue). Cependant, les tests d'homogénéité sans série de référence sont beaucoup moins fiables et ne devraient jamais être effectués avec une procédure entièrement automatique. Il convient d'inspecter visuellement les séries candidates originales et désaisonnalisées et d'étudier toutes les métadonnées disponibles afin d'établir la liste finale des points de changement à ajuster. L'estimation de l'ampleur de l'ajustement à apporter est également plus incertaine. La plus grande incertitude de ces jeux de données homogénéisés devrait être quantifiée et clairement communiquée.

Une raison importante pour laquelle l'homogénéisation absolue est moins précise est que la série est plus «bruyante». En outre, ce «bruit» est partiellement dû aux variations à long terme du système climatique et est plus difficile à distinguer des inhomogénéités que le bruit non corrélé.

On peut limiter les problèmes en repérant et modélisant les variations à basse fréquence dans le cadre d'une procédure d'homogénéisation, par exemple en utilisant la méthode développée par Wen *et al.* (2011). Un exemple d'application de cette méthode à la série chronologique de Fort Nelson (Canada) sur la nébulosité est présenté à la figure 5. Cette méthode a permis de recenser correctement deux écarts, ainsi qu'un cycle de 12,5 ans, un cycle annuel et une tendance négative; la combinaison de ces composantes est représentée par la courbe bleue du graphique. Si le cycle de 12,5 ans de cette série chronologique était ignoré, c'est-à-dire si l'algorithme PMFred était appliqué directement à cette série chronologique, trois faux points de changement seraient décelés et il ne serait pas possible de déterminer le point de changement réel. Ces techniques de réduction du bruit fonctionnent particulièrement bien pour l'homogénéisation absolue, mais elles peuvent également être explorées pour l'homogénéisation relative.

Comme cela est mentionné ci-dessus, les séries de référence peuvent être dérivées d'autres variables. Par exemple, Wan *et al.* (2010) et Minola *et al.* (2016) ont utilisé comme séries de référence les vitesses du vent géostrophique dérivées des gradients de pression au niveau de la mer sur une zone triangulaire (formée par trois stations d'observation de la pression de surface) afin d'homogénéiser les données de vitesse du vent de surface provenant des stations situées dans cette zone. Cependant, le vent géostrophique n'est probablement pas une bonne référence pour les vents de surface sur les régions tropicales/subtropicales et les régions à topographie complexe. Dai *et al.* (2011) ont utilisé les relations empiriques entre les anomalies de la température de l'air et de la tension de vapeur dérivées d'observations récentes, lorsque des rapports de dépression du point de rosée étaient disponibles dans ces conditions, pour ajuster les effets

d'échantillonnage artificiel en estimant les rapports de dépression du point de rosée manquants pour les conditions froides ( $T < 30$  °C) et sèches (dépression du point de rosée artificiellement fixée à 30 °C). Pour les stations côtières et insulaires, les températures de surface de la mer peuvent constituer une autre série de référence possible.

La réanalyse est indépendante des données de surface pour la plupart des variables et peut donc servir de série de référence lorsqu'il n'existe aucune autre série appropriée. Cette approche a été utilisée pour l'homogénéisation des données relatives au vent dans la péninsule ibérique et en Australie (Azorin-Molina *et al.*, 2014 et 2019). Pour les données d'observation en altitude, des données de réanalyse sont utilisées dans les jeux de données RAOBCORE et RICH (Haimberger *et al.*, 2012).

Comme les températures en altitude (sur lesquelles se fondent les réanalyses) ont généralement des échelles de décorrélation plus longues que les températures de surface, les réanalyses peuvent servir de série de référence lorsqu'il n'existe pas de bonnes données de surface provenant de stations voisines. Pour la même raison, on s'attend à ce que les inhomogénéités de la réanalyse soient visibles dans les séries chronologiques de différences avec de nombreuses stations, ce qui pourrait permettre à l'opérateur de déterminer si l'inhomogénéité se trouve dans les données de la station ou dans la réanalyse. Des ruptures d'homogénéités de grande échelle figurent dans les données de réanalyse lorsque l'on introduit ou modifie des jeux de données satellitaires. Les modèles atmosphériques présentent souvent des différences (régionales) au niveau des observations. Lorsque davantage d'observations sont disponibles, la réanalyse peut passer (progressivement) de l'état de base du modèle à celui de climatologie observée.

Si les données contiennent une rupture nette dont la taille ne peut être estimée avec une précision suffisante, l'homogénéisation absolue peut également être utilisée pour déterminer la date de cette rupture et pour ne pas tenir compte des données antérieures à cette date limite. Cependant, dans le cas de régions où les données sont rares, comme l'Arctique, le fait de ne pas utiliser ces données peut créer un biais. Dans le cas des données des périodes antérieures, la numérisation d'un plus grand nombre de données peut aider à résoudre ce conflit.

## 2.6 Ajustements statistiques

La façon la plus précise de calculer les corrections est de les considérer simultanément pour toutes les ruptures d'un réseau (régional); ce principe est appelé «correction conjointe». Cette méthode a été introduite par Caussinus et Mestre (2004): elle décompose les données brutes en un signal climatique régional pour toutes les stations et une fonction en escalier pour modéliser les ruptures et le bruit de chaque station, lequel est minimisé. Elle a permis d'améliorer les corrections pour presque toutes les contributions à l'étude comparative HOME<sup>1</sup> (Venema *et al.*, 2012) qui n'intégraient pas encore cette méthode (Domonkos *et al.*, 2013). Lorsque toutes les ruptures sont détectées, cette méthode élimine en général parfaitement le biais de tendance à grande échelle, tandis que son incertitude est déterminée par le bruit de la série chronologique de différences; les erreurs dans l'ensemble des ruptures détectées entraînent une sous-corrrection de toute erreur de tendance (Lindau et Venema, 2018b).

De nombreuses méthodes d'homogénéisation ne reconnaissent pas deux points de rupture proches l'un de l'autre. Il est courant de traiter deux ruptures proches comme une seule rupture dans la correction, en utilisant les données qui précèdent la première rupture et celles qui suivent la deuxième rupture comme base des ajustements à plus long terme, et de traiter séparément les données qui se trouvent entre deux ruptures. À titre d'exemple, selon la méthodologie australienne ACORN-SAT, les ruptures ne sont traitées séparément que si elles sont espacées d'au moins quatre ans (Trewin, 2018), et les ruptures détectées à l'aide du progiciel RHtests avec (ou sans) série de référence sont espacées d'au moins 5 (ou 10) points de données.

Ces méthodes de correction peuvent être appliquées aux données annuelles, saisonnières et mensuelles. De nombreuses inhomogénéités ont un cycle saisonnier et ne seraient pas corrigées si l'on se contentait de calculer des corrections à l'échelle annuelle et de les appliquer comme des corrections fixes à chaque mois. Il ressort des expériences numériques menées avec la méthode d'homogénéisation PRODIGE sur le jeu de données de référence HOME que les corrections mensuelles de température pour chaque mois civil distinct étaient très précises. En raison du cycle saisonnier des inhomogénéités, les corrections annuelles ont donné de moins bons résultats. Pour les précipitations, les corrections annuelles étaient très précises, bien que les inhomogénéités aient un cycle saisonnier. En revanche, l'incertitude des corrections

<sup>1</sup> Action pour l'homogénéisation de l'initiative COST (Coopération européenne en science et technologie)

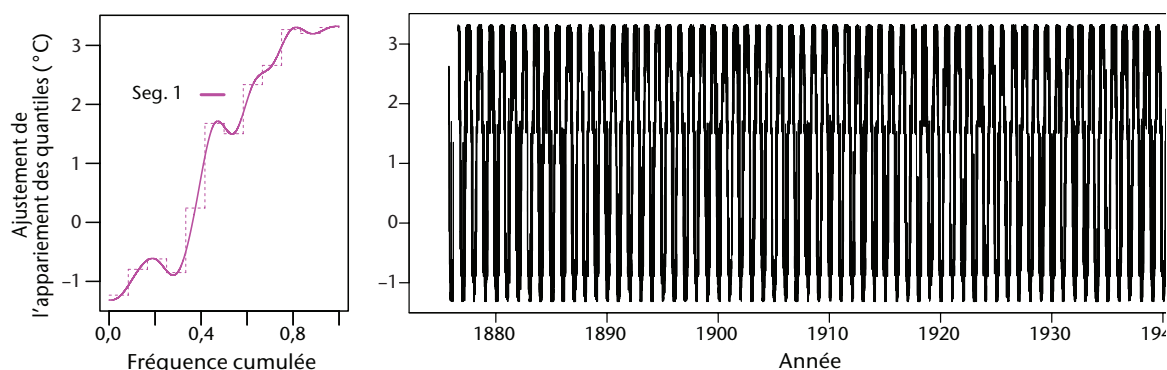
mensuelles estimées pour chaque mois civil distinct était probablement trop importante pour les données de précipitations, ce qui a conduit à une homogénéisation moins précise. Corriger les températures à une échelle mensuelle/saisonnière et les précipitations à une échelle annuelle est probablement une bonne règle empirique, bien que dans les réseaux plus clairsemés que les réseaux européens habituels, étudiés dans HOME, les corrections de température pourraient se comporter de la même manière que les corrections de précipitations. L'élément le plus important est probablement le SNR plutôt que l'élément météorologique.

Les méthodes de correction appliquées à la distribution des données journalières (Trewin et Trevitt, 1996; Della-Marta et Wanner, 2006; Mestre *et al.*, 2011; Wang *et al.*, 2010; Wang *et al.*, 2013; Trewin, 2013) peuvent également être utilisées pour corriger les données ayant un cycle saisonnier, qui dominent généralement la variance totale. Par exemple, lorsqu'une station canadienne située près du Saint-Laurent a été déplacée vers un site à l'intérieur des terres, l'inhomogénéité de la série des températures quotidiennes maximales de l'air en surface a clairement montré un cycle saisonnier (voir la figure 6, image de droite). Le site à l'intérieur des terres est beaucoup plus chaud en été et un peu plus froid en hiver que le site proche du fleuve.

Une question pratique à régler est de savoir s'il convient d'utiliser l'ensemble de la sous-période homogène du relevé des données pour estimer les ajustements. Par défaut, la période entière est utilisée pour exploiter pleinement la quantité limitée de données afin d'estimer les corrections, mais il peut parfois être judicieux de s'écarter de ce principe. Lorsque la sous-période homogène est longue, l'avantage supplémentaire d'un plus grand nombre de données diminue, tandis qu'augmente le risque d'inhomogénéités restantes ou de différences notables de changement climatique entre la candidate et la référence. Cela est particulièrement vrai lorsque la référence n'est pas optimale. Lorsque l'homogénéisation absolue est appliquée, il est courant de limiter les périodes à 10 ans avant et après la rupture. Il est préférable de ne pas utiliser de données (ajustées) au-delà des sous-périodes homogènes adjacentes.

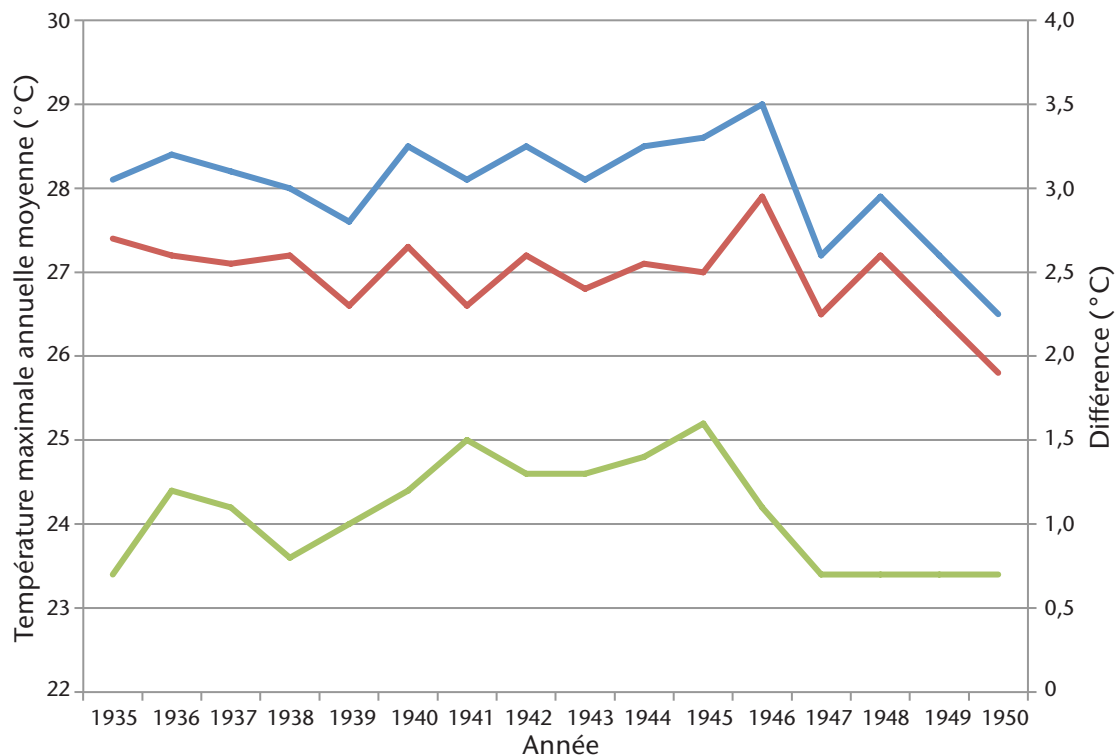
Un autre problème, évoqué plus haut, se présente lorsque la période précédant ou suivant immédiatement une inhomogénéité n'est pas représentative du comportement à long terme de la station. Cette situation peut se produire, par exemple, lorsqu'une station est déplacée après une détérioration soudaine de son exposition. Dans de tels cas, il serait approprié d'exclure les données enregistrées entre la modification de l'exposition et le déplacement du site lors d'ajustements à plus long terme (voir la figure 7). En pratique, les problèmes de ce type sont souvent difficiles à détecter dans les tests statistiques. De même, lorsque la date de la rupture est incertaine (SNR inférieur à un; Lindau et Venema, 2016) ou que les ruptures statistiques et les métadonnées suggèrent des dates différentes, il peut être justifié d'exclure certaines données relevées autour de la rupture lors de l'estimation des corrections.

Pour le calcul des corrections, il n'est pas recommandé d'utiliser des composites sans supprimer les séries comportant des ruptures. Les ruptures qui biaisent le plus les changements de la



**Figure 6.** Effet, sur les températures maximales quotidiennes de l'air en surface, du déplacement d'une station depuis un site proche du fleuve vers un site à l'intérieur des terres. Il s'agit de la distribution en quantiles (image de gauche) et de la série temporelle (image de droite) des ajustements de l'appariement des quantiles estimés à partir d'une série de différences (candidate moins référence) qui sont nécessaires pour ajuster les données du site proche du fleuve (extrait de la figure 1 de Wang *et al.* (2013)).





**Figure 7. Un exemple de données non représentatives avant un changement. À Gayndah (Australie) (ligne bleue; axe de gauche), l'abri s'est progressivement détérioré à partir de 1940, avant d'être remplacé en octobre 1945. La différence de température maximale (ligne verte; axe de droite) entre Gayndah et la moyenne de trois sites de référence – Dalby, Brisbane et Emerald (ligne rouge; axe de gauche), est passée de 1,0 °C à environ 1,5 °C dans les années précédant le changement d'abri, avant de retomber à 0,7 °C après ce changement.**

moyenne du réseau sont celles qui apparaissent dans toutes les stations. Lorsque cela se produit sur une courte période, la référence aura un biais similaire à la candidate, et le biais à grande échelle subsistera en grande partie après correction.

Il est d'usage de corriger les données pour qu'elles correspondent aux conditions de la section homogène la plus récente. Ainsi, les données entrantes futures seront toujours homogènes, à moins que d'autres changements n'interviennent à la station.

## 2.7 Examen des données et cycles d'homogénéisation multiples

L'étape finale est la validation des données homogénéisées. Quelle que soit la qualité de l'homogénéisation des données, la perfection ne sera pas atteinte et une certaine inhomogénéité résiduelle subsistera dans les séries ajustées. Il est nécessaire de procéder à une évaluation critique des travaux menés et de revoir les données homogénéisées.

Cet examen devrait porter sur chaque série prise séparément et consister à évaluer si les nouvelles valeurs sont logiques: le cycle saisonnier est-il préservé? Les valeurs se situent-elles dans la fourchette attendue pour la station? Les ajustements varient-ils radicalement entre des mois adjacents? etc. Si un jeu de données complet a été homogénéisé, il est extrêmement utile d'examiner la cohérence régionale de l'évolution temporelle des séries, ainsi que de comparer les séries d'ajustement avec les changements connus intervenus dans le réseau et de calculer les ajustements effectués pour des types d'inhomogénéités spécifiques. Il peut s'agir de réaliser des tests d'homogénéité sur les données homogénéisées – par exemple, en les comparant avec celles d'autres stations homogénéisées de la région, ou en recherchant des tendances anormales dans telle ou telle station homogénéisée. Si le réseau comprend plusieurs types de stations, il peut être utile de les comparer. On peut comparer les résultats obtenus avec ceux d'exercices

d'homogénéisation précédents et de pays voisins. La cohérence entre les éléments climatiques devrait également être étudiée, par exemple, la température moyenne, maximale et minimale. Outre la quantification des résultats, l'inspection visuelle des ajustements, des données homogénéisées et des séries de différences peut également aider à détecter les problèmes.

Dans le cas d'une homogénéisation manuelle avec des méthodes fondées sur une référence composite, il est normal qu'un processus d'homogénéisation initial ne permette pas de résoudre tous les problèmes d'homogénéité du jeu de données considéré. Plusieurs raisons peuvent expliquer cette situation. Parmi les plus courantes, on notera:

- Des inhomogénéités non détectées dans une ou plusieurs séries de référence ou dans une paire d'observations parallèles à l'une des stations pendant la période d'observations parallèles;
- Des conditions climatiques anormales au moment d'une rupture d'homogénéité, entraînant un ajustement non représentatif (par exemple, une période particulièrement humide ou sèche immédiatement avant ou après une rupture d'homogénéité);
- Des conditions sur un site, peu avant ou après un ajustement, qui ne sont pas représentatives du relevé à plus long terme. Un scénario courant est celui d'un déplacement en raison de travaux de construction récents; les données de l'ancien site relevées après le début des travaux peuvent ne pas être représentatives de la période antérieure et ne devraient donc pas être utilisées pour déterminer l'ajustement nécessaire pour un jeu de données à long terme.

Des anomalies peuvent également survenir pour d'autres raisons; par exemple, un problème de qualité des données décelé à une station au cours d'un mois donné pendant une période d'observations parallèles peut, selon la méthode utilisée, avoir des répercussions sur les ajustements de ce mois mais pas sur d'autres.

La manière la plus efficace de traiter ces questions est d'effectuer un deuxième cycle d'homogénéisation en conjonction avec un examen des données.

Lorsque des problèmes sont recensés au cours du deuxième cycle, les options pour les résoudre dépendent de leur nature et peuvent consister à:

- Répéter l'homogénéisation en utilisant un jeu différent de stations de référence, si possible. Une approche utile, si l'on se sert d'un certain nombre de séries de référence, consiste à estimer l'ampleur de l'ajustement qui serait généré en utilisant une seule station de référence, pour chaque station prise séparément, et à supprimer toute station de référence qui génère des résultats excessivement anormaux par rapport aux autres stations de référence;
- Utiliser une période différente comme base d'ajustement; par exemple, si une station a déménagé en 1951 mais qu'un nouveau bâtiment a été construit près de l'ancienne station en 1949, il conviendrait d'utiliser la période se terminant en 1948 (plutôt que la période se terminant en 1950 ou 1951) comme base d'ajustement à long terme.

Lorsque l'on effectue plusieurs cycles d'homogénéisation, il ne faudrait jamais supposer que les données d'un cycle précédent étaient homogènes, mais calculer plutôt à nouveau toutes les corrections. Sinon, le résultat risque de s'éloigner de la vérité en raison d'une homogénéisation répétée.

Parfois, même après un deuxième (ou troisième) cycle d'homogénéisation, certaines stations afficheront encore des tendances anormales par rapport à d'autres stations. Cette situation peut résulter de changements qui affectent progressivement le climat local sur une période prolongée (par exemple, le site est envahi par une zone urbaine en expansion ou l'agriculture irriguée s'étend de plus en plus dans le district), ou bien d'effets locaux naturels (par exemple, l'augmentation des niveaux de remontée d'eau côtière dans l'océan refroidit un site côtier). S'il peut être établi avec un niveau de confiance raisonnable que ces tendances anormales ont une cause spécifique non climatique

(telle que l'urbanisation), on peut par exemple les supprimer du jeu de données ou les exclure de certains produits fondés sur le jeu de données (par exemple, en n'incluant pas les stations influencées par l'urbanisation dans les évaluations du changement climatique à long terme).

## 2.8 Documentation

Lorsqu'un jeu de données ajustées est élaboré, les ajustements devraient être correctement étayés et publiés. Il conviendrait notamment d'indiquer les dates couvertes par l'ajustement, le moyen de détection de l'inhomogénéité (méthodes statistiques ou métadonnées) et, si possible, la cause probable de l'ajustement. Des statistiques sommaires sur l'influence des ajustements (globaux et par catégorie de métadonnées) peuvent aider l'utilisateur à évaluer la qualité des données. Une bonne pratique consiste à partager les données brutes et les données homogènes ainsi que les métadonnées relatives aux ruptures décelées.

Les méthodes utilisées pour élaborer chacun des jeux de données homogénéisés devraient être correctement expliquées sous une forme accessible. Dans l'idéal, il s'agirait d'un article en libre accès publié dans la littérature scientifique et évalué par des pairs. Ce document devrait contenir au moins une description détaillée des méthodes utilisées, disponible au même endroit que les données homogénéisées elles-mêmes. La meilleure pratique consiste à rédiger un code clair et bien étayé, en gardant à l'esprit qu'il sera publié en même temps que les données.

## 2.9 Maintenance opérationnelle d'un jeu de données homogénéisé

Le développement initial d'un jeu de données homogénéisé est une entreprise importante. Normalement, l'un des principaux objectifs des jeux de données homogénéisés est de servir de jeux de données sous-jacents pour les produits (par exemple, une anomalie de température nationale ou mondiale), ce qui signifie que pour continuer d'actualiser des produits, il faut aussi continuer de mettre à jour les jeux de données sous-jacents correspondants. L'un des avantages des méthodes d'homogénéisation automatique est qu'elles peuvent être facilement appliquées à de nouvelles données.

Les jeux de données homogénéisés ont tendance à être construits de telle manière que les données les plus récentes ne sont pas ajustées. Cela permet d'actualiser le jeu de données en y ajoutant de nouvelles données sans autre ajustement. Une exception peut être faite lorsque les données plus anciennes sont considérées comme une référence plus fiable à long terme. Par exemple, dans un réseau d'observation des précipitations essentiellement manuel mais dans lequel un petit nombre de stations automatiques sont mises en place, il peut être plus approprié d'ajuster les stations automatiques pour que leurs données soient équivalentes aux données manuelles antérieures et que l'on obtienne ainsi une meilleure cohérence spatiale de l'ensemble du réseau. Une autre raison d'ajuster les observations manuelles antérieures est qu'elles sont souvent plus précises et de meilleure qualité que les observations automatisées.

Avec le temps, un jeu de données homogénéisé deviendra obsolète. Deux facteurs majeurs y contribuent. Tout d'abord, certaines stations qui font partie du jeu de données initial fermeront avec le temps (elles seront parfois remplacées dans le réseau par de nouvelles stations voisines, qui peuvent servir de base pour établir une série composite). Ensuite, de nouvelles inhomogénéités peuvent se produire dans les stations qui restent dans le jeu de données.

Il est recommandé de procéder, au moins tous les cinq ans, à une réévaluation de tout jeu de données homogénéisé. Cette réévaluation devrait inclure les activités suivantes:

- Vérifier l'état de toutes les stations du jeu de données existant et, si des stations sont fermées ou ne communiquent plus de manière fiable, vérifier s'il est possible de les remplacer par une autre station qui peut alors être intégrée dans un relevé composite;
- Chercher les métadonnées récentes (couvrant la période écoulée depuis la dernière mise à jour) de toutes les stations du jeu de données;
- Intégrer toutes les données anciennes pertinentes qui sont devenues disponibles (par exemple, les données récemment numérisées dans le cadre d'activités de sauvetage de données);

- Dans le cas de méthodes manuelles, il conviendrait d'effectuer au moins un test statistique pour détecter les inhomogénéités dans la partie la plus récente du relevé. Ce test comprend une réévaluation portant sur les dernières années de la version précédente du jeu de données, car les inhomogénéités des dernières (ou des premières) années d'une série chronologique sont difficiles à détecter et à quantifier et les nouvelles observations supplémentaires peuvent permettre d'effectuer des évaluations plus fiables. Il est recommandé, en particulier pour les méthodes automatiques, d'homogénéiser l'ensemble du jeu de données.

## 2.10 **Problèmes à l'échelle du réseau et solutions pour les résoudre**

Il arrive que des changements aient des répercussions sur l'ensemble des stations d'un réseau national ou sur un grand nombre d'entre elles, de façon simultanée ou sur une période de plusieurs années. On peut notamment citer, à titre d'exemples:

- Une modification de l'heure d'observation, soit explicite (comme le changement de la période d'observation pour les données quotidiennes de 00.00-00.00 UTC à 06.00-06.00 UTC au Canada en 1961), soit implicite (par exemple, les stations australiennes ont continué à effectuer des observations à la même heure locale lorsque l'heure d'été a été mise en place au début des années 1970, ce qui a introduit un décalage réel d'une heure de l'heure d'observation standard pendant l'été);
- Un changement majeur de type d'instrument, tel qu'un changement d'abri de thermomètre standard (par exemple, lors de l'automatisation des stations météorologiques), ou l'introduction d'un nouveau type de radiosonde (les observations en altitude sont particulièrement sensibles à ce type de changement, car elles ne reposent guère sur des infrastructures fixes, d'où la possibilité de mettre en œuvre des changements assez rapidement);
- Un changement dans les procédures d'observation ou les définitions, comme le passage de l'unité de mesure de la nébulosité de 1/10 à 1/8, ou un changement d'unités (par exemple, de Fahrenheit à Celsius);
- Un changement dans les algorithmes utilisés pour l'analyse des données – comme un changement dans la définition de la température moyenne quotidienne (par exemple, le passage de la moyenne de huit observations de trois heures à la moyenne des températures maximale et minimale quotidiennes).

Les changements à l'échelle du réseau peuvent être particulièrement difficiles à gérer dans un processus d'homogénéisation. Comme ils s'appliquent habituellement à la plupart ou à la totalité des stations d'une région donnée, l'utilisation de stations de référence du même réseau sera d'une utilité limitée tant pour détecter l'inhomogénéité que pour déterminer son impact probable. En outre, un changement dont l'impact peut ne pas être significatif ou détectable dans une station donnée (par exemple, une inhomogénéité de température de 0,2 °C) peut être significatif dans une moyenne nationale, s'il concerne toutes les stations d'un réseau, ou dans la moyenne mondiale, s'il représente un changement technologique ou organisationnel généralisé.

Voici quelques stratégies possibles pour faire face à ces changements:

- Les changements affectant l'ensemble du réseau ne sont pas toujours mentionnés dans les bases de métadonnées, qui présentent généralement les changements connus intervenant dans telle ou telle station. En revanche, comme il s'agit d'événements nationaux importants, ils sont souvent mentionnés dans les rapports annuels;
- Si un changement affecte la plupart des stations d'un réseau mais pas toutes (par exemple, l'ensemble des stations automatiques, mais pas les stations manuelles), comparer les stations concernées aux autres;

- Comparer les observations avec celles qui ont été effectuées près de la frontière dans les pays voisins qui ne sont pas concernés par le changement (cette méthode n'est efficace que s'il existe de telles stations; elle suppose normalement une frontière terrestre et un accès aux observations des autres pays, ce qui n'est pas toujours facile à obtenir);
- Comparer les données avec un autre type de données qui n'a pas été affecté par le changement en question. Par exemple, comparer les températures à la surface avec les températures de radiosondage à 850 hPa (ou avec les réanalyses reposant sur celles-ci), ou comparer les vitesses de vent mesurées avec les vitesses du vent géostrophique dérivées des champs de pression au niveau moyen de la mer;
- Utiliser des données alternatives pour indiquer l'impact possible du changement. Par exemple, pour un changement ancien de l'heure d'observation, même si des données infraquotidiennes haute résolution de la période où le changement a eu lieu ne sont pas disponibles, il peut être possible d'utiliser des données haute résolution des dernières années pour estimer l'impact d'un changement ancien de l'heure d'observation. Par exemple, Vincent *et al.* (2009 et 2012) ont utilisé des données de température horaires pour corriger un biais dans les températures minimales quotidiennes causé par un changement de l'heure d'observation au Canada. En Autriche (Böhm *et al.*, 2010) et en Espagne (Brunet *et al.*, 2006), des expériences parallèles ont été réalisées pour quantifier les changements dus au passage aux abris Stevenson.

Ces méthodes peuvent produire des résultats assez grossiers, et ne permettent de quantifier les répercussions de ces inhomogénéités (ou de déterminer qu'un changement n'a pas eu d'impact significatif) qu'au niveau national ou régional, sans tenir pleinement compte des différents impacts qu'une inhomogénéité a pu avoir sur telle ou telle station.

Les problèmes d'inhomogénéité à l'échelle du réseau peuvent être résolus de différentes manières en fonction des ressources disponibles. La première version du jeu de données du Réseau de données climatologiques anciennes des États-Unis contenait un ajustement physique pour la transition des abris Stevenson (*Cotton Region shelters*) aux systèmes MMTS (systèmes de températures maximales et minimales) de stations météorologiques automatiques. Cette transition s'est produite dans de nombreuses stations sur une courte période et les données ont donc été difficiles à homogénéiser par homogénéisation statistique, en comparant une station candidate à une référence composite (signal moyen sur plusieurs stations voisines), car les stations voisines étaient souvent également concernées. Les ajustements se sont fondés sur les estimations des stations candidates dont les stations voisines n'avaient pas connu cette transition. Après avoir conçu une nouvelle méthode d'homogénéisation, en faisant des comparaisons par paires, il est devenu possible de surmonter cette difficulté et les ajustements physiques n'ont plus été appliqués.

## 2.11 Défis spécifiques aux jeux de données multinationaux

La combinaison de données provenant de différents réseaux et pays présente des avantages car elle réduit les problèmes d'inhomogénéité à l'échelle du réseau. Toutefois, la mise au point d'un jeu de données homogènes à l'échelle mondiale ou régionale (multinationale) présente également certains défis, notamment les suivants:

- **Jeux de données de grande taille.** Ces jeux de données contiennent souvent des informations provenant de centaines ou de milliers de stations. Ils dépassent en général la taille limite sur le plan pratique pour les méthodes d'homogénéisation manuelles et nécessitent le recours à des méthodes automatisées ou semi-automatisées.
- **Accès limité aux métadonnées.** Les métadonnées sont habituellement archivées au niveau national et seules les métadonnées les plus basiques sont échangées au niveau international, raison pour laquelle, il n'est guère possible d'intégrer les métadonnées dans le processus d'homogénéisation des jeux de données multinationaux (même si les métadonnées peuvent être consultées, leur utilisation nécessitera souvent l'interprétation de documents rédigés dans la langue locale). Parfois, il peut même être difficile de

déterminer exactement de quelle station les données proviennent et une des tâches les plus importantes de l'élaboration d'un jeu de données mondiales consiste à repérer et harmoniser les jeux de données en double provenant de différentes sources (par exemple, Rennie *et al.*, 2014). Il est à espérer que l'Outil d'analyse de la capacité des systèmes d'observation (OSCAR) permettra d'améliorer le partage des métadonnées.

- **Accès limité aux séries de référence potentielles et à d'autres données pertinentes.** Les jeux de données multinationaux ne sont généralement constitués que de données provenant de stations climatologiques sélectionnées. En revanche, dans les jeux de données nationaux, il est souvent possible de s'appuyer sur des données supplémentaires (par exemple, des données de stations exploitées sur une plus courte période et dont les données sont en nombre insuffisant pour être intégrées dans un jeu de données à long terme mais qui peuvent néanmoins être utiles comme stations de référence pour des sous-périodes spécifiques), ainsi que sur des données d'éléments météorologiques supplémentaires et des données infraquotidiennes, alors que les jeux de données mondiaux sont généralement à élément unique. Le Centre européen pour les prévisions météorologiques à moyen terme (CEPMMT) travaille sur une base de données multi-éléments dans laquelle les observations de plusieurs éléments météorologiques d'une même station sont conservées ensemble (Dunn et Thorne, 2017).

Par conséquent, des méthodes d'homogénéisation automatiques sont utilisées pour ces jeux de données, et les corrélations entre les séries candidates et de référence seront plus faibles, ce qui rendra plus difficile la détection des petites inhomogénéités et augmentera l'incertitude des ajustements qui sont effectués.

L'une des solutions consiste à s'appuyer, dans la mesure du possible, sur des informations de portée nationale. Dans le jeu de données HadCRUT sur la température mondiale, géré par l'université d'East Anglia et le Service météorologique britannique, des jeux de données homogénéisés au niveau national sont utilisés lorsqu'ils sont disponibles (Jones *et al.*, 2012). Xu *et al.* (2017) se sont appuyés sur plusieurs jeux de données nationaux homogénéisés, qu'ils ont élargis pour élaborer un jeu de données mondial homogénéisé.

## 2.12 **Conclusion: bonnes pratiques en matière d'homogénéisation**

Le chapitre 2 a passé en revue un ensemble de problèmes et de considérations en rapport avec la mise en place de jeux de données homogènes portant sur de longues périodes. La mesure dans laquelle ces considérations peuvent être mises à profit varie considérablement en fonction des possibilités qu'ont les développeurs d'avoir accès aux données et métadonnées pertinentes, des outils et systèmes informatiques dont ils disposent, du soutien qu'ils peuvent obtenir, de la densité du réseau d'observation sous-jacent et de la taille du jeu de données considéré.

Il existe un certain nombre de principes qui peuvent être considérés comme de bonnes pratiques pour l'homogénéisation des données. On peut citer notamment les suivants:

1. C'est en associant méthodes statistiques et métadonnées que l'homogénéisation des données est la plus efficace (par exemple, Yosef *et al.*, 2018). Si cela n'est pas possible (par exemple, parce que les métadonnées ne sont pas disponibles ou parce que l'absence de séries de référence rend l'homogénéisation statistique difficile), l'homogénéisation sera probablement moins efficace;
2. Il conviendrait de toujours procéder à une homogénéisation statistique; on ne peut pas supposer que les métadonnées sont parfaites;
3. Si un jeu de données présente des problèmes connus (par exemple, une modification de l'heure d'observation à l'échelle du réseau), ces derniers devraient être résolus avant d'envisager des inhomogénéités plus spécifiques à telle ou telle station. En général, l'utilisation de séries de référence directement ou par comparaison par paires ne permet pas de repérer les changements à l'échelle du réseau;
4. Les séries de référence devraient être utilisées lors de l'application des méthodes d'homogénéisation statistique, si cela est possible;

5. Il est important d'essayer d'obtenir un SNR supérieur à un et il est utile de maximiser encore le SNR;
6. Une fois qu'une version préliminaire d'un jeu de données homogénéisé est obtenue, les méthodes qui se fondent sur l'hypothèse que la référence composite est homogène nécessitent de procéder à un deuxième cycle d'homogénéisation, comme cela est indiqué dans la section 2.7 ci-dessus;
7. Les jeux de données homogénéisés devraient être entièrement mis à jour au moins tous les cinq ans;
8. Lorsque des changements majeurs sont prévus, il conviendrait de procéder à des observations parallèles pendant au moins deux ans. Brandsma *et al.* (2019) présentent des renseignements utiles sur la gestion des changements dans les réseaux;
9. Lorsqu'un jeu de données ajustées est élaboré, les ajustements devraient être correctement étayés et publiés;
10. Les méthodes utilisées pour élaborer chacun des jeux de données homogénéisés devraient être correctement présentées sous une forme accessible.

Bien qu'aucune méthode ne puisse garantir l'homogénéité parfaite d'un jeu de données, et que tout jeu de données homogénéisé présente un certain niveau d'incertitude lié aux ajustements nécessaires à sa création, le respect de ces principes devrait porter à son maximum la probabilité qu'un jeu de données soit suffisamment homogène pour être utilisé efficacement aux fins du développement de produits climatologiques portant sur le long terme.

---

## CHAPITRE 3. CHOISIR UN LOGICIEL D'HOMOGENÉISATION STATISTIQUE

Le présent chapitre a pour objet de guider le lecteur à travers les méthodes numériques et les logiciels qui peuvent être utilisés pour diverses opérations d'homogénéisation.

### 3.1 Méthodes d'homogénéisation statistique

La liste et le tableau ci-dessous décrivent (dans l'ordre alphabétique) les logiciels d'homogénéisation accessibles à tous, qui étaient utilisés en climatologie au moment de la rédaction de la présente publication. De nouvelles méthodes étant régulièrement mises au point, la liste ci-dessous ne devrait pas être considérée comme exhaustive. D'autres méthodes sont décrites dans la littérature scientifique mais ne sont pas mentionnées ici car les logiciels correspondants n'ont pas encore été mis en circulation sous une forme utilisable par la communauté au sens large. Au moment de la rédaction de la présente publication, le tableau était mis à jour à l'adresse suivante: <http://www.climatol.eu/tt-hom/>.

Bon nombre des descriptions données ci-dessous renvoient à une étude comparative des méthodes d'homogénéisation menée dans le cadre du projet HOME (Venema *et al.*, 2012). La section 3.2.2 contient des informations plus détaillées sur cette étude.

**ACMANT** est une méthode d'homogénéisation des données de température et de précipitation. C'est l'une des méthodes automatiques les plus précises qui existe pour l'homogénéisation de séries de données de température sans métadonnées.

**AnClim** est un outil qui met en application toutes les méthodes de détection et de correction courantes avec une interface utilisateur graphique. Utilisant une approche d'ensemble comprenant de nombreuses méthodes et de multiples paramètres, AnClim n'a pas contribué de façon marquante à l'étude comparative HOME. Cela étant, ce logiciel permet d'avoir accès à de nombreuses méthodes qui peuvent également être utilisées indépendamment, de manière plus standard. AnClim, ainsi que ProClimDB (non gratuit), contribuent à l'automatisation de nombreuses tâches liées aux bases de données.

**Bayesian MDL** est une méthode de détection de points de rupture multiples qui est disponible gratuitement mais qui en est encore au stade de la recherche. Elle est mentionnée par soucis d'exhaustivité par les concepteurs de méthodes d'homogénéisation (Li *et al.*, ArXiv 2017).

**Berkeley Earth** est une méthode d'homogénéisation et d'interpolation utilisée pour les jeux de données de température mondiaux. Les corrections d'homogénéisation sont calculées dans la partie interpolation. Ce ne sont pas les données de la station qui sont renvoyées, mais un champ ou une estimation du climat régional à l'emplacement de la station.

**Climatol** applique le test SNHT (*Standard Normal Homogeneity Test*) pour diviser la série en sous-périodes homogènes. Il calcule une série complète pour chaque sous-période homogène. Dans la dernière étape du processus, toutes les données manquantes sont estimées à partir d'autres sous-périodes de la même station (lorsqu'elles sont disponibles) ou de stations voisines. Climatol est l'un des logiciels qui tolère le mieux les lacunes dans les données et qui peut tirer parti des métadonnées disponibles.

**GAHMDI** résout le problème des points de rupture multiples grâce à un algorithme de recherche global (algorithme génétique). Il est proposé avec HOMAD, une méthode permettant de corriger la distribution des données journalières. Aucune comparaison ou documentation n'est disponible en dehors de deux articles publiés (Toreti *et al.*, 2010 et 2012).

**GSIMCLI** utilise des méthodes géostatistiques pour calculer numériquement l'hypothèse nulle en recourant à l'approche de Monte-Carlo. Il s'agit d'une nouvelle méthode qui n'a fait l'objet d'aucune étude comparative aussi ne dispose-t-on d'aucune information sur ses performances. Elle fournit une interface utilisateur graphique et permet une approche automatisée pour les réseaux.



**HOMER** a été conçu dans le cadre de l'Action COST «HOME»<sup>1</sup>, mais ses performances n'ont pas été évaluées. Il met en œuvre plusieurs méthodes de détection des points de rupture multiples. Utilisant l'option par paires, il s'inscrit dans le prolongement de la méthode de détection des points de rupture multiples PRODIGE et devrait donc constituer l'une des meilleures méthodes manuelles. Il inclut la méthode de correction ANOVA, qui est probablement la méthode de correction la plus précise qui existe. Il est préférable de ne pas utiliser l'option détection conjointe (logiciel R "cghseg" mis en œuvre dans HOMER) toute seule (voir Gubler *et al.*, 2017).

**iCraddock** applique le test Craddock par paires. Il s'agit d'une méthode manuelle subjective mais qui fonctionne bien pour les petits réseaux et qui est recommandée dans l'étude comparative HOME. Cette méthode peut également être utilisée pour les données journalières (Brugnara *et al.*, 2012).

**MASH** est un algorithme d'homogénéisation automatique qui se base sur des tests d'hypothèse. Il est conçu pour fonctionner avec des références non homogènes et utilise une méthode de détection des sauts multiples. C'est un excellent outil.

**PHA**, ou algorithme d'homogénéisation par paires, est utilisé par l'Administration américaine pour les océans et l'atmosphère (NOAA) pour homogénéiser ses jeux de données de température à l'échelle nationale (*U.S. Historical Climatology Network* (USHCN)) et mondiale (Réseau mondial de données climatologiques anciennes (GHCN)). C'est un outil très fiable et recommandé pour les grands ensembles de données. Il peut utiliser des métadonnées.

**ReDistribution Test** est un test sur les points de rupture pris séparément pour le vecteur vent.

**RHtests** met en œuvre plusieurs tests de détection de ruptures en tenant compte des autocorrélations et de la distance des marges. RHtests et AnClim sont les seules méthodes de la liste qui peuvent homogénéiser des données sans utiliser des séries de référence. Lorsqu'elle est utilisée, la série de référence doit être calculée par l'opérateur. RHtests comprend un test pour les points de rupture documentés. Il convient de noter que ces tests ne sont pas conçus pour être utilisés avec une procédure entièrement automatique; il est nécessaire d'effectuer une analyse des résultats de la détection automatisée pour finaliser les résultats.

**SNHT** renvoie à un logiciel R dans le réseau CRAN (*Comprehensive R Archive Network*), qui met en œuvre le test bien connu *Standard Normal Homogeneity Test* mais d'une manière modifiée par Haimberger (2007) et selon l'approche par paires utilisée par Menne et Williams (2009).

---

<sup>1</sup> Coopération européenne en science et technologie (COST); un programme de financement pour stimuler la collaboration en Europe.

Tableau 1. Présentation générale des caractéristiques des logiciels d'homogénéisation

| Logiciel            | Résolution (détection) <sup>a</sup> | Méthode de détection        | Utilisation de références <sup>b</sup> | Résolution (correction) <sup>c</sup>  | Méthode de correction                           | Opération principale       | Utilisation de métadonnées | Variable <sup>d</sup> | Documentation <sup>e</sup>          | Référence  |
|---------------------|-------------------------------------|-----------------------------|--|---------------------------------------|---|----------------------------|----------------------------|-----------------------|-------------------------------------|--|
| ACMANT <sup>f</sup> | Année, mois                         | Points de rupture multiples | Composite                              | Année, mois, [jour]                   | Conjointe (ANOVA)                               | Automatique                | Non                        | Toute                 | Guide d'utilisation                 | Domonkos et Coll (2017)                              |
| AnClim              | Année, mois                         | Plusieurs                   | Composite, par paires                  | Année, mois, jour                     | Plusieurs                                       | Interactive, automatique   | Oui                        | Toute                 | Manuels                             | Štěpánek <i>et al.</i> (2009)                        |
| Berkeley Earth      | Mois                                | Division                    | Composite                              | s/o                                   | s/o   | Automatique                | Oui                        | T                     | Article                             | Rohde <i>et al.</i> (2013)                           |
| Climatol            | Mois (série), jour                  | Division                    | Composite                              | Mois (série), jour                    | Comblement des données manquantes               | Automatique                | Oui                        | Toute                 | Manuel et guide d'utilisation       | Guijarro (2018)                                      |
| GAHMDI HOMAD        | Mois (série), jour                  | Points de rupture multiples | Sélection                              | Jour                                  | Méthode des moments d'ordre supérieur           | Automatique                | Oui                        | T                     | Aucune                              | Toreti <i>et al.</i> (2010, 2012)                    |
| GSIMCLI             | Année, mois                         | Points de rupture multiples | Composite                              | Voir note de bas de page <sup>9</sup> | Voir note de bas de page <sup>9</sup>           | Automatique et interactive | Non                        | T, p                  | Manuels                             | Ribeiro <i>et al.</i> (2017), Costa et Soares (2009) |
| HOMER               | Année, saison, mois                 | Points de rupture multiples | Par paires, conjointe                  | Année, mois                           | Conjointe (ANOVA)                               | Interactive                | Oui                        | Toute                 | Guide d'utilisation de base + cours | Mestre <i>et al.</i> (2013)                          |
| iCraddock           | Année, saison, mois                 | Division                    | Par paires                             | Année, saison, mois, jour             | Quotidiennement: corrections mensuelles lissées | Interactive                | Oui                        | Toute                 | Aucune                              | Craddock (1979), Brunetti <i>et al.</i> , 2006       |
| MASH                | Année, saison, mois                 | Points de rupture multiples | Composite                              | Mois, [jour]                          | Comparaisons multiples                          | Automatique et interactive | Oui                        | Toute                 | Guide d'utilisation                 | Szentimrey (2008, 2014)                              |
| ReDistribution Test | Relevés                             | Points de rupture uniques   | Pas de référence                       | s/o                                   | s/o   | Interactive                | Non (mais interactif)      | Vent                  | Aucun                               | Petrovic (2004)                                      |
| RHtests             | Année, mois, jour                   | Division                    | Sélection ou pas de référence          | Année, mois, jour                     | Régression multiphase                           | Interactive                | Oui                        | Toute                 | Guide d'utilisation + cours         | Wang (2008a et b), Wang et Feng (2013)               |

| Logiciel        | Résolution (détection) <sup>a</sup> | Méthode de détection | Utilisation de références <sup>b</sup> | Résolution (correction) <sup>c</sup> | Méthode de correction               | Opération principale | Utilisation de métadonnées | Variable <sup>d</sup> | Documentation <sup>e</sup> | Référence                                    |
|-----------------|-------------------------------------|----------------------|--|--------------------------------------|-------------------------------------|----------------------|----------------------------|-----------------------|----------------------------|--|
| Logiciel R SNHT | Année, mois                         | Division             | Composite et par paires                | Mois                                 | Composite et comparaisons multiples | Automatique          | Non                        | T                     | Fichiers d'aide            | Haimberger (2007), Menne et Williams, (2009) |
| PHA             | Année                               | Division             | Par paires                             | Année, [mois]                        | Comparaisons multiples              | Automatique          | Oui                        | T                     | Notes en texte clair       | Menne <i>et al.</i> (2009)                   |

<sup>a</sup> Sauf indication contraire dans la colonne Résolution, "mois" signifie que la détection porte sur de multiples séries mensuelles en parallèle;

<sup>b</sup> Les options sont les suivantes: sélection de la référence de l'opérateur, calcul de la moyenne (référence composite), suppression des références avec des ruptures dans le composite (pas de référence), détection par paires et détection conjointe;

<sup>c</sup> Les crochets signifient que la résolution est prise en compte par le logiciel, mais que les corrections ne sont pas calculées à cette résolution;

<sup>d</sup> Les options sont les suivantes: T = température; p = précipitations; Toute = distribution gaussienne et log-normale ou modèles additifs et multiplicatifs;

<sup>e</sup> Un guide d'utilisation est limité à quelques pages et est plus court qu'un manuel;

<sup>f</sup> ACMANT peut détecter des ruptures dans les moyennes annuelles et le cycle saisonnier en parallèle;

<sup>g</sup> Les corrections sont calculées au niveau de la résolution de la donnée (séries annuelles ou mensuelles). Les corrections sont appliquées à une mesure, calculée selon la méthode GSIMCLI (définie par l'utilisateur: percentile, moyenne ou médiane), de la fonction de densité de probabilité (pdf) de la station candidate, qui est estimée à l'aide de références composites.

### 3.2 Efficacité des méthodes d'homogénéisation statistique

Il existe deux indications permettant de savoir si l'homogénéisation a amélioré un jeu de données ou pas. Premièrement, les ruptures relevées doivent correspondre aux ruptures constatées dans l'historique de la station. Deuxièmement, si le jeu de données était à nouveau homogénéisé, il ne présenterait aucune autre rupture et les résultats devraient être cohérents au niveau régional, cohérents en eux-mêmes et plausibles du point de vue climatologique (voir la section 2.7 sur l'examen des données). Toutefois, ces évaluations n'empêchent pas d'établir un trop grand nombre de ruptures et de supprimer ainsi la variabilité climatique régionale réelle (sur-homogénéisation). En outre, ces indications ne sont pas assez précises pour permettre de choisir la meilleure méthode d'homogénéisation statistique. Par conséquent, pour choisir les méthodes d'homogénéisation appropriées, on s'appuie essentiellement sur ce que l'on sait de leur efficacité d'après les informations que l'on trouve dans la littérature scientifique.

Il existe deux sources de données sur les performances générales des méthodes d'homogénéisation statistique: les principes théoriques (section 3.2.1) et les études numériques (section 3.2.2). Ces deux sources d'information se renforcent mutuellement et sont toutes deux nécessaires pour donner confiance dans les méthodes proposées. Le principe sur lequel repose la conception d'une méthode d'homogénéisation peut être solide sur le plan théorique, mais les caractéristiques de mise en œuvre ont aussi leur importance, et la méthode peut se révéler peu performante lors d'une comparaison numérique. Inversement, les études numériques ne font que tester des scénarios spécifiques, qui peuvent ne pas être applicables pour la tâche à accomplir, et les informations dont on dispose nous aident à déterminer ce qui est important et réalisable.

#### 3.2.1 Principes théoriques

Si une série de différences non corrélées (série candidate moins série de référence) ayant une distribution normale contient une rupture à une date connue, le test approprié est un simple test t pour la différence de la moyenne avant et après la rupture. Si la même série ne contient qu'une seule rupture mais que sa position est inconnue, il faut envisager d'effectuer de multiples contrôles, et le test qui convient dans ce cas de figure est le test SNHT (*Standard Normal homogeneity Test*) (Alexandersson, 1986) ou le test PMT (*Penalized Maximal t Test*) (Wang *et al.*, 2007).

Cependant, les séries de données climatologiques contiennent généralement plus d'une rupture, et la série de référence peut elle aussi contenir des ruptures, qui ne devraient pas être faussement attribuées à la série à traiter. Il semble que ce qui détermine essentiellement la performance des méthodes d'homogénéisation statistique est la manière dont elles résolvent ces deux problèmes.

Il existe trois façons de détecter de multiples points de rupture dans une même série:

1. Parfois, des tests de rupture unique sont effectués sur fenêtres mobiles. Toutefois, pour réduire la probabilité qu'il y ait des points de rupture multiples dans la fenêtre, celle-ci doit être étroite. Cela rend la méthode moins sensible et cette approche n'est pas très utilisée;<sup>2</sup>
2. Traditionnellement, on utilise des méthodes conçues pour détecter un seul point de rupture: la série est divisée au point de rupture le plus significatif, après quoi les deux nouvelles séries sont testées à nouveau (division hiérarchique et ses variantes);
3. Les méthodes modernes de détection des ruptures multiples testent efficacement toutes les combinaisons possibles de ruptures multiples.

<sup>2</sup> La méthode de détection sur fenêtres mobiles peut être utile lorsque l'on veut supprimer uniquement les ruptures nettes, et non les inhomogénéités progressives. Par exemple, pour étudier le réchauffement progressif (non linéaire) dû à l'urbanisation, on peut vouloir supprimer uniquement l'effet des relocalisations (Zhang *et al.*, 2014).

En théorie, les méthodes de détection des points de rupture multiples constituent le moyen le plus précis de résoudre ce problème.

Il existe également plusieurs approches pour traiter les inhomogénéités dans les séries de référence:

1. Calculer la moyenne des valeurs d'un grand nombre de stations de référence. Cela permet de supprimer les grands sauts évidents dans la série de référence composite, mais il arrive souvent qu'une grande partie d'un réseau connaisse une transition similaire sur quelques années ou décennies. Le biais dû à cette transition se retrouverait également en grande partie dans une telle référence composite et pourrait réduire les capacités de détection;
2. Sélectionner des stations de référence qui ne présentent pas de discontinuité autour du point de rupture dans la station candidate ou bien corriger les ruptures dans les séries de référence avant de les utiliser. Ces approches doivent être utilisées de manière itérative, car les ruptures dans les séries de référence doivent être relevées. L'utilisation de données préalablement homogénéisées n'est pas sans risque. Par conséquent, ces approches doivent être validées avec un soin particulier au cas où elles éliminent des biais à grande échelle;
3. Détecter les ruptures au niveau de paires de stations. Selon cette approche, la station de référence n'est pas supposée être homogène, les ruptures dans les deux stations sont détectées comme des ruptures dans les séries de différences entre elles. Une deuxième étape «attribution» est nécessaire pour déterminer à quelle station appartiennent les ruptures détectées dans les paires;
4. Procéder à une détection conjointe et simultanée de toutes les ruptures dans un réseau de multiples stations. Il s'agit d'un puzzle combinatoire compliqué, qui exige plus de calculs que les autres méthodes.

La détection conjointe est théoriquement la meilleure solution. Toutefois, cette méthode n'est utilisée pour l'instant que dans le logiciel d'homogénéisation HOMER, et les résultats ne sont pas concluants. En pratique, on préfère donc les méthodes 2 et 3.

Une nouvelle méthode de correction conjointe (souvent appelée ANOVA) a été mise au point par Caussinus et Mestre (2004). Cette méthode décompose un réseau de stations en un signal climatique régional, une fonction en escalier applicable pour chaque station afin de modéliser les inhomogénéités et le bruit. Les corrections sont calculées en minimisant le bruit. Les corrections effectuées avec cette méthode sont non biaisées si toutes les ruptures sont correctement identifiées (Lindau et Venema, 2018b). Ainsi, toutes les ruptures devraient être incluses dans le processus d'ajustement, y compris celles qui sont proches les unes des autres. Il a été démontré que cette méthode donnait des résultats plus précis que les méthodes traditionnelles utilisées pour les réseaux européens denses simulés dans HOME (Domonkos *et al.*, 2013).

### 3.2.2 **Études numériques**

Traditionnellement, les études de validation ont été axées sur les résultats des méthodes de détection des ruptures. Cela peut aider à comprendre le fonctionnement de l'algorithme, mais on ne voit pas clairement quel est le compromis optimal pour l'analyse climatologique des données homogénéisées entre le taux de succès et le taux de fausses alertes. Des travaux plus récents ont inclus des mesures de l'erreur, telles que l'erreur quadratique moyenne et l'incertitude restante dans la tendance après homogénéisation, qui évaluent la performance de méthodes d'homogénéisation complètes et présentent un intérêt pour l'utilisateur des données dans le domaine du climat (Domonkos, 2011; Venema *et al.*, 2012; Williams *et al.*, 2012).

Dans une certaine mesure, les résultats des études de validation dépendront du ou des indicateur(s) utilisé(s) pour l'évaluation. La manière dont l'homogénéisation est effectuée peut également influencer les résultats (par exemple, des métadonnées ont-elles été utilisées, les algorithmes ont-ils été exploités et l'opérateur a-t-il reçu une bonne formation). Dans le cas des méthodes manuelles et semi-automatiques, des différences nettes ont été constatées entre

les opérateurs (Venema *et al.*, 2012), aussi convient-il de faire une distinction nette entre la méthode/le logiciel d'homogénéisation et les ensembles de données homogénéisées validées (contributions). S'agissant en particulier des RHtests, qui mettent en œuvre plusieurs tests de détection et méthodes de correction et peuvent être utilisés de nombreuses façons, les résultats d'une étude de validation peuvent ne pas être représentatifs.

Williams *et al.* (2012) n'ont étudié que l'algorithme d'homogénéisation par paires, Domonkos (2011) a comparé un grand nombre d'algorithmes d'homogénéisation automatique, tandis que l'Action COST «HOME» (Venema *et al.*, 2012) couvre quasiment toutes les méthodes les plus récentes et les plus utilisées, y compris plusieurs méthodes manuelles. Il convient de garder à l'esprit que ces trois études ont été réalisées pour des réseaux denses, la performance des méthodes sera donc moindre pour les réseaux plus clairsemés. En outre, la classification des méthodes pourrait être différente pour d'autres réseaux.

Ces résultats numériques confortent l'idée selon laquelle les algorithmes conçus pour résoudre les problèmes liés aux points de rupture multiples et à l'inhomogénéité des séries de référence permettent d'obtenir des résultats beaucoup plus précis que les méthodes traditionnelles. L'étude comparative HOME a recommandé les algorithmes suivants : ACMANT, iCraddock, MASH et PRODIGE et, pour les grands réseaux, PHA. Là encore, il convient de noter que si l'étude avait été réalisée de façon différente avec un autre jeu de données de référence, elle aurait pu aboutir à une recommandation/conclusion différente. Les résultats de l'étude HOME, en particulier, ne sont pas représentatifs de l'ensemble du logiciel RHtestsV3. En fait, aucune étude de référence appliquant des méthodes selon une procédure entièrement automatique ne peut évaluer l'ensemble des RHtests car ce logiciel a été conçu pour un fonctionnement interactif, comprenant des interventions manuelles ou semi-automatiques.

Pour mener à bien les études de validation ci-dessus il a fallu créer un réseau fictif de stations dont les inhomogénéités étaient connues. Le jeu de données de l'étude comparative HOME visait à modéliser les réseaux de stations de mesure de la température et des précipitations pour l'Europe. Ses données de validation sont assez réalistes, mais la variance des ruptures est environ deux fois trop élevée. L'étude n'incluait pas de biais de tendance à grande échelle clairement définis; les biais étaient donc petits et difficiles à corriger. En outre, la forte densité de stations que l'on note dans l'étude n'est pas représentative des débuts de la période instrumentale ou des réseaux clairsemés comme ceux des pays d'Europe du Nord ou des pays en développement. Une étude comparative relevant de l'Initiative internationale sur les températures de surface (ISTI) est en cours (Thorne *et al.*, 2011); elle vise à résoudre ces problèmes (Willett *et al.*, 2014). De nouveaux résultats de validation étaient enregistrés dans le cadre du projet MULTITEST (<http://www.climatol.eu/MULTITEST/>) au moment de la rédaction des présentes directives.

Dans l'étude comparative HOME, toutes les contributions ont rendu les données de température plus homogènes, sauf celle utilisant l'homogénéisation absolue. Cela montre à quel point les méthodes absolues peuvent être dangereuses, notamment lorsqu'elles sont utilisées dans le cadre d'une procédure entièrement automatique comme dans l'étude comparative (voir section 2.5). Il convient cependant de noter que pour obtenir une comparaison à l'aveugle, le signal climatique régional utilisé dans le jeu de données HOME était plus variable que la normale et qu'il était donc plus difficile à homogénéiser qu'un jeu de données d'observation réel. Néanmoins, l'homogénéisation absolue est généralement moins fiable et devrait être utilisée avec plus de précaution et jamais automatiquement. S'agissant des données de précipitation, seules les meilleures méthodes ont pu améliorer l'homogénéité.

Les études de validation peuvent également être basées sur des données homogénéisées de haute qualité. Par exemple, dans Gubler *et al.* (2017), des informations sur les inhomogénéités basées sur des données homogénéisées provenant d'un réseau de stations très dense ont été utilisées pour étudier la performance des méthodes d'homogénéisation appliquées à un réseau plus clairsemé. L'avantage de cette procédure est que les inhomogénéités correspondent par définition à la réalité. Par contre, même si la densité du réseau de stations est élevée, l'homogénéisation ne sera pas parfaite. Ce genre d'étude présente une synergie avec celles qui utilisent des données simulées. Gubler *et al.* (2017) ont étudié quatre façons différentes d'utiliser le logiciel HOMER et ont découvert que pour la détection conjointe, il valait mieux ne pas utiliser HOMER tout seul. La manière dont les métadonnées ont été utilisées dans cette étude n'a pas amélioré les résultats. Les ruptures dues à la transition vers les stations météorologiques automatiques n'ont souvent pas été détectées dans le réseau clairsemé, alors que ces transitions sont censées être le point fort de ces méthodes par paires.

Kuglitsch *et al* (2012) ont utilisé des données suisses homogénéisées pour valider les méthodes d'homogénéisation PRODIGE et Toreti et la fonction FindU.wRef de RHtestsV3. Les résultats indiquent que PRODIGE détecte davantage de ruptures qui peuvent être confirmées par des métadonnées, mais qu'il a également un taux de fausses alertes bien plus élevé que RHtestsV3 ou la méthode Toreti: PRODIGE a trouvé 1 140 ruptures (dont 515 sont confirmées par des métadonnées, soit 45,2 %), tandis que RHtestsV3 et la méthode Toreti ont trouvé respectivement 438 et 683 ruptures, dont 72,4% et 70,3% sont confirmées par des métadonnées. Cependant, l'étude a porté uniquement sur la détection des ruptures; elle n'a pas évalué les méthodes d'ajustement, ni l'ensemble de la procédure d'homogénéisation (c'est-à-dire la détection et l'ajustement combinés). L'effet de la surestimation du nombre de ruptures ou de l'absence de détection de ruptures réelles sur les résultats de l'homogénéisation n'est donc pas clair.

### 3.3 Méthodes automatisées et manuelles

Les méthodes d'homogénéisation peuvent être entièrement automatisées, et l'on parle alors de méthodes automatiques (c'est-à-dire de méthodes qui ne requièrent aucune autre intervention humaine que la sélection des jeux de données), ou bien faire appel à quelques opérations manuelles, et être appelées méthodes manuelles. Les méthodes manuelles peuvent nécessiter une intervention humaine pour, entre autre:

- Choisir les stations de référence à utiliser (alors qu'une méthode automatique peut utiliser, par exemple, un critère basé purement sur la distance ou la corrélation);
- Fusionner des informations provenant des méthodes statistiques et des métadonnées;
- Déterminer quelles inhomogénéités identifiées par une méthode statistique devraient être retenues;
- Déterminer sur quelle période devrait porter la comparaison (pour comparer deux sites, on n'utilisera pas, par exemple, une année marquée par des anomalies climatiques ou une période où les données sont moins fiables).

Les méthodes automatiques et manuelles ont été utilisées avec succès dans de nombreux pays. Les méthodes manuelles ont l'avantage de permettre d'introduire des informations sur une station qui peuvent ne pas être facilement quantifiables (par exemple, des changements progressifs connus au niveau de l'environnement du site), et de relever plus facilement des résultats anormaux obtenus dans des stations individuelles (bien que le risque de résultats anormaux soit réduit si certaines des pratiques décrites dans la section 2.7 sur l'examen des données sont mises en œuvre). De nombreuses méthodes manuelles peuvent être entièrement automatisées en remplaçant l'intervention humaine par une procédure automatique correspondante.

Toutefois, les méthodes manuelles présentent l'inconvénient d'être à forte intensité de main-d'œuvre et de nécessiter l'avis d'un expert, qui peut ne pas être disponible dans certains cas. En outre, on peut mieux valider les méthodes automatiques car il est facile de calculer de nombreux cas de figure et paramètres, et on peut ainsi améliorer plus rapidement leurs capacités. Cela signifie aussi que l'on dispose d'estimations plus fiables des incertitudes. En outre, en supprimant les incertitudes liées aux interventions humaines, on a plus de chance d'obtenir la précision voulue dans les études de validation.

Les méthodes automatisées ou semi-automatisées sont recommandées dans les cas où une prise de décision n'est pas nécessaire (voir les quatre points ci-dessus) et pour les opérateurs ayant une expérience limitée en matière d'homogénéisation. Elles constituent également la seule option pratique pour les très grands jeux de données, en particulier les jeux de données mondiaux ou régionaux.

### 3.4 Cas d'utilisation

La performance des méthodes d'homogénéisation mentionnées dans la section précédente est un critère important à prendre en considération, qui aura clairement un impact sur la qualité de vos données homogénéisées, même si la plupart des méthodes énumérées dans le présent chapitre amélioreront au moins les données de température dans la plupart des circonstances. On trouvera dans la présente section toute une gamme de cas d'utilisation qui montrent comment pondérer divers critères et choisir un logiciel qui corresponde à une tâche donnée.

Un autre critère à prendre en considération est celui de la capacité du logiciel de traiter de grandes quantités de métadonnées. La taille du réseau est importante: plus il est grand, plus les méthodes automatiques sont à privilégier. Sa densité est également un facteur important: il est plus facile d'homogénéiser les données provenant de réseaux à haute densité que celles de réseaux à moyenne ou basse densité.

En outre, le choix de recourir à une méthode spécifique peut tenir au fait que l'on dispose d'experts au niveau local ou de possibilités de formation; l'étude comparative HOME a mis en évidence des différences nettes entre les contributions des différents opérateurs utilisant la même méthode, ce qui peut être lié à l'expérience acquise. Les méthodes automatiques sont moins tributaires des services d'experts et sont recommandées pour les utilisateurs moins expérimentés. Mais les méthodes d'homogénéisation absolue ne doivent jamais être appliquées automatiquement.

Une fois que l'importance des ruptures documentées (ruptures à des moments connus) a été déterminée, le logiciel RHtests peut être utilisé: c'est le seul outil qui peut tester en tandem les ruptures documentées et non documentées. Si un seul point de rupture dans une série chronologique de différences à un moment donné doit être vérifié, le t-test mis en œuvre dans de nombreux logiciels de calcul scientifique peut être utilisé.

Lorsque la densité du réseau est trop faible et qu'il n'y a de fait aucune série de référence utilisable (séries de données de réanalyse ou indirectes, notamment), la seule option est de recourir à l'homogénéisation absolue. Cette approche est mise en œuvre dans les logiciels RHtests et AnClim.

Si l'on est en présence d'un petit réseau de moins de 50-100 stations avec peu de lacunes dans les données, et que l'on connaît un peu l'historique des stations et dispose de suffisamment de temps pour obtenir un jeu de données de qualité, la méthode par paires HOMER et iCraddock sont de bonnes options. Les avantages d'HOMER par rapport à iCraddock tiennent au fait que l'emplacement des ruptures dans les paires est déterminé objectivement, ce qui accélère aussi la procédure, et HOMER prend automatiquement en charge la correction conjointe. Si le réseau est peu dense et que les références peuvent avoir des signaux climatiques légèrement différents, iCraddock est une bonne option car l'opérateur peut évaluer graphiquement si des différences sont à prévoir sur le plan climatologique ou s'il y a lieu de suspecter une inhomogénéité. Lorsque les options interactives ne sont pas utilisées, ACMANT et MASH peuvent également fonctionner avec de petits réseaux. Dans ce cas, l'opérateur sélectionne soigneusement les séries à utiliser pour l'homogénéisation. Pour les très grands réseaux (qui comportent des centaines ou des milliers de stations), cette approche devient trop compliquée. La taille du réseau est une considération relativement flexible: le développeur de iCraddock a homogénéisé un jeu de données avec environ 700 séries sur la base de son expérience.

Pour les réseaux de taille moyenne (plus de 100 stations), les méthodes automatiques ACMANT, Climatol et MASH sont des options intéressantes. MASH peut gérer jusqu'à 500 stations, ACMANT v.4 jusqu'à 5 000, et la taille du réseau de Climatol n'est limitée que par la mémoire de l'ordinateur (ce qui représente dans la pratique quelques milliers de stations). Climatol et MASH (si le volume de données manquantes n'est pas trop important) seraient préférables si l'on veut tirer parti des métadonnées disponibles.

L'algorithme d'homogénéisation par paires a été conçu pour les jeux de données continentaux et mondiaux et peut également traiter des jeux de données perturbés avec des séries courtes et des données manquantes. Tels seraient les principaux éléments à prendre en considération si l'on veut opter pour cette méthode. Avec une certaine automatisation des interventions humaines, le RHtests est la seule autre méthode qui a été utilisée pour homogénéiser des jeux de données mondiaux (Xu *et al.*, 2017).



## CHAPITRE 4. HISTORIQUE DE L'HOMOGENÉISATION

On connaît depuis longtemps l'incidence que peut avoir le manque d'homogénéité des données sur leur analyse et les origines de l'homogénéisation remontent loin. En septembre 1873, lors du Congrès météorologique international de Vienne, Carl Jelinek exprimait la nécessité de disposer d'informations sur les séries de données nationales pluriannuelles (k.k. (kaiserlich-königlich) Hof- und Staatsdruckerei, 1873). Des décennies plus tard, en 1905, lors de la conférence internationale des directeurs de services météorologiques, G. Hellmann (k.k. Zentralanstalt für Meteorologie und Geodynamik, 1906) regrettait encore l'absence de séries chronologiques de données climatologiques homogènes en raison des modifications de l'environnement des stations et du remplacement des instruments, et plaidait pour que les stations dont les relevés portent sur de longues périodes, les Säkularstationen (stations d'observation centenaires), soient maintenues aussi homogènes que possible. Bien que les participants à cette conférence aient recommandé de veiller à ce que les conditions d'exploitation d'un nombre suffisant de stations ne soient pas modifiées, ce problème fondamental de manque d'homogénéité se pose toujours aujourd'hui.

### 4.1 Détection et ajustement

Dans les premiers temps, on a supprimé les points de changement documentés à l'aide de mesures parallèles. Compte tenu des heures d'observation différentes à l'observatoire astronomique de la k.k. Université de Vienne (Autriche), les données ont été ajustées en utilisant les mesures pluriannuelles sur 24 heures de l'observatoire astronomique de la k.k. Université de Prague (aujourd'hui République tchèque). Les mesures de Milan (Italie), entre 1763 et 1834, ont été ajustées à des moyennes sur 24 heures en utilisant les mesures de Padoue (Kreil, 1854a et b).

Au début du XX<sup>e</sup> siècle, Conrad (1925) a appliqué et évalué le critère de Heidke (Heidke, 1923) en utilisant les rapports de deux séries de précipitations. Suite à cela, il a recommandé d'utiliser des critères supplémentaires pour tester l'homogénéité des séries avec une succession et une alternance de signes algébriques: le critère de Helmert (Helmert, 1907) et le critère minutieux d'Abbe (Conrad et Schreier, 1927). Le critère d'Helmert pour les paires de stations et le critère d'Abbe étaient encore considérés comme des outils appropriés dans les années 1940 (Conrad 1944). Quelques années plus tard, la méthode de la double masse a été vulgarisée pour la détection des ruptures (Kohler, 1949).

### 4.2 Séries de référence

Julius Hann (1880) a étudié la variabilité des quantités et des rapports absolus de précipitations entre les stations. Il a utilisé ces rapports pour le contrôle de la qualité. Brückner (1890) s'est inspiré de cette approche pour rechercher des inhomogénéités dans les données de précipitation en comparant des stations voisines; il n'a utilisé aucune statistique.

Dans leur ouvrage *Methods in Climatology*, Conrad et Pollak (1950) ont formalisé cette technique d'homogénéisation relative, qui est maintenant la principale méthode utilisée pour détecter et éliminer les effets des changements artificiels. L'élaboration de séries de référence, consistant à moyenner les données de nombreuses stations situées dans une zone géographique relativement petite, a par la suite été recommandée par le Groupe de travail de l'OMM sur les fluctuations climatiques (OMM, 1966).

Les articles d'Alexandersson (1986) et d'Alexandersson et Moberg (1997) ont fait largement connaître le test SNHT (*Standard Normal Homogeneity Test*). Ce test, qui s'accompagnait d'indications claires sur la manière de l'utiliser ainsi que de références pour homogénéiser les données des stations, a été largement adopté.

### 4.3 **Développements récents**

Le SNHT est une méthode conçue pour détecter un point de rupture unique, or les séries climatiques contiennent en général plus d'une rupture. Aussi la mise au point de méthodes spécialement conçues pour détecter et corriger de multiples points d'inflexion et travailler avec des références non homogènes (Szentimrey, 1999; Mestre, 1999; Caussinus et Mestre, 2004) a t'elle constitué un grand pas en avant. L'étude comparative de l'Action COST «HOME» de l'Union européenne (Venema et al., 2012) a montré que les méthodes de ce type étaient plus précises.

Un article de Caussinus et Mestre (2004) a également décrit pour la première fois une méthode qui corrige simultanément toutes les séries d'un réseau. Cette méthode de correction conjointe a amélioré la précision de toutes les contributions à l'étude comparative HOME, sauf une qui n'utilisait pas encore cette approche (Domonkos et al., 2013).

Les travaux en cours visant à créer des jeux de données appropriés pour l'étude de la variabilité du climat et du changement climatique ont favorisé le développement continu de meilleures méthodes de détection et de correction des points d'inflexion. Pour renforcer ce processus, le Service météorologique hongrois (OMSZ) a lancé en 1996 une série de séminaires sur l'homogénéisation (OMSZ, 1996; OMM, 1999; OMSZ, 2001; OMM, 2004; OMM, 2006; OMM, 2010; OMM, 2011; OMM, 2014; OMSZ, 2017).

---

## CHAPITRE 5. CONTEXTE THÉORIQUE DE L'HOMOGENÉISATION

Le présent chapitre aborde certains aspects théoriques de l'homogénéisation des séries chronologiques mensuelles et réexamine nombre des problèmes d'homogénéisation mentionnés dans les chapitres précédents. Il s'adresse à tous ceux qui souhaitent acquérir une meilleure compréhension analytique et aux scientifiques désireux d'élaborer leurs propres méthodes.

Dans la pratique, les séries mensuelles sont, pour la plupart, homogénéisées sous forme de moyennes. Le processus d'homogénéisation vise à détecter les défauts d'homogénéité d'une moyenne et à ajuster la série.

Pour déceler les défauts d'homogénéité d'une série mensuelle et procéder à un ajustement, il convient de résoudre les problèmes mathématiques suivants :

- Modélisation statistique spatio-temporelle de la série;
- Méthode de comparaison de la série à traiter et de la série de référence;
- Point de rupture (point de changement) et détection des valeurs aberrantes;
- Méthode d'ajustement des séries.

Le présent chapitre portera aussi sur les points suivants :

- Procédures de contrôle de la qualité;
- Rétablissement des données manquantes;
- Utilisation des métadonnées;
- Méthodes manuelles et méthodes automatiques;
- Évaluation des méthodes (théoriques et de référence).

### 5.1 Structure générale des modèles spatio-temporels additifs

La modélisation statistique spatio-temporelle d'une série est une question fondamentale. Les comparaisons à opérer, la détection du point de rupture et les méthodes d'ajustement dépendent du modèle statistique retenu. Si les séries de données sont distribuées de manière gaussienne (par exemple dans le cas de la température), on peut utiliser un modèle spatio-temporel additif. Appliqué aux séries mensuelles provenant de plusieurs stations d'une même région climatique (de faible étendue), ce modèle additif prend la forme suivante:

$$X_{j,m}(t) = \mu_m(t) + S_{j,m} + IH_{j,m}(t) + \varepsilon_{j,m}(t), \quad (j = 1, 2, \dots, N; m = 1, 2, \dots, 12; t = 1, 2, \dots, n), \quad (1)$$

où

|                       |   |
|-----------------------|---|
| $j = 1, 2, \dots, N$  | Indicatif de la station   |
| $m = 1, 2, \dots, 12$ | Indicatif du mois   |
| $t = 1, 2, \dots, n$  | Année   |
| $\mu_m(t)$            | Signal commun et inconnu de changement climatique, valeurs temporelles attendues ou évolution temporelle concernant les stations  |
| $S_{j,m}$             | Valeurs spatiales attendues ou évolution spatiale concernant les stations   |
| $IH_{j,m}(t)$         | Signaux d'inhomogénéité s'apparentant à une « variation d'un échelon », en supposant généralement des points de rupture inconnus $T$ et des modifications $IH_{j,m}(T) - IH_{j,m}(T+1) \neq 0$ , et $IH_{j,m}(n) = 0$ . |

Par conséquent, les valeurs ou moyennes attendues sont les suivantes:

$$E(X_{j,m}(t)) = \mu_m(t) + S_{j,m} + IH_{j,m}(t), \quad (j = 1, 2, \dots, N; m = 1, 2, \dots, 12; t = 1, 2, \dots, n),$$

et superposées à ces moyennes se trouvent les séries de bruit gaussien, qui prennent la forme vectorielle suivante:

$$\boldsymbol{\varepsilon}_m(t) = [\varepsilon_{1,m}(t), \dots, \varepsilon_{N,m}(t)]^T \sim N(\mathbf{0}, C_m), \quad (m = 1, 2, \dots, 12; t = 1, 2, \dots, n).$$

Les matrices  $C_m$  ( $m = 1, 2, \dots, 12$ ) comprennent les covariances spatiales entre les stations et elles présupposent l'absence de changement climatique ou de rupture d'homogénéité au fil des ans. Les méthodes actuelles d'homogénéisation des séries mensuelles présupposent en effet qu'il n'y a pas de changement climatique ou de défaut d'homogénéité dans les moments d'ordre plus haut.

En général, les termes du vecteur bruit,  $\boldsymbol{\varepsilon}_m(t)$ , ont également des autocorrélations temporelles, comme on le voit ci-dessous (Remarque N° 1, points 1.1 et 1.2).

L'homogénéisation a pour but de détecter le signal d'inhomogénéité  $IH_{j,m}(t)$  et d'ajuster la série mensuelle brute  $X_{j,m}(t)$  comme suit:

$$X_{H,j,m}(t) = X_{j,m}(t) - I\hat{H}_{j,m}(t) \quad (j = 1, 2, \dots, N; m = 1, 2, \dots, 12; t = 1, 2, \dots, n),$$

où  $I\hat{H}_{j,m}(t)$  est le signal d'inhomogénéité estimé.

#### Remarque N° 1

Dans la pratique, les méthodes absolues et les méthodes relatives présentent quelques différences.

Homogénéisation absolue: on utilise les séries de données d'une seule station:  $N = 1$ . Le principal inconvénient des méthodes absolues est qu'il est quasiment impossible de distinguer le signal de changement climatique et de variabilité  $\mu_m(t)$ , lorsqu'il n'est pas connu, de la rupture d'homogénéité  $IH_{j,m}(t)$  si l'on ne dispose pas d'un complément d'information sur la variabilité du signal climatique, auquel cas des données homogènes seraient nécessaires.

Homogénéisation relative: on utilise des données de plusieurs stations –  $N > 1$  – et on les compare les unes avec les autres, ce qui permet d'éliminer le signal commun de changement climatique et de variabilité du climat  $\mu_m(t)$ , qu'on ne connaît pas.

Pour résoudre le problème que pose l'homogénéisation, il existe deux approches fondamentales:

- 1.1 Les séries mensuelles (voir l'équation 1 ci-dessus) sont examinées à la suite les unes des autres, une par une, par ordre chronologique. Ce type d'approche soulève les questions suivantes:
  - Le cycle annuel ou la saisonnalité de  $\mu_m(t)$ ,  $E_{j,m}$ ,  $IH_{j,m}(t)$ ,  $C_m$  ( $m = 1, 2, \dots, 12$ ); les dernières covariances tiennent compte de façon implicite des écarts types et des corrélations spatiales;
  - L'autocorrélation temporelle entre les éléments de mois adjacents.
- 1.2 Les séries mensuelles, saisonnières et annuelles peuvent être examinées en parallèle, indépendamment les unes des autres. Ce peuvent être les séries suivantes:
  - Les séries mensuelles afférentes à chaque mois du calendrier ( $m = 1, 2, \dots, 12$ ), considérées séparément:  

$$X_{j,m}(t) = \mu_m(t) + S_{j,m} + IH_{j,m}(t) + \varepsilon_{j,m}(t) \quad (j = 1, 2, \dots, N; t = 1, 2, \dots, n)$$
  - Les séries saisonnières afférentes à chaque saison du calendrier ( $s = 1, 2, 3, 4$ ), considérées séparément:  

$$X_{j,s}(t) = \mu_s(t) + S_{j,s} + IH_{j,s}(t) + \varepsilon_{j,s}(t) \quad (j = 1, 2, \dots, N; t = 1, 2, \dots, n)$$
  - Les séries annuelles:  

$$X_{j,y}(t) = \mu_y(t) + S_{j,y} + IH_{j,y}(t) + \varepsilon_{j,y}(t) \quad (j = 1, 2, \dots, N; t = 1, 2, \dots, n).$$

Pour ce type d'approche, il n'est pas nécessaire d'examiner le cycle annuel ou la saisonnalité ainsi que l'autocorrélation temporelle. Dans le cas d'un mois, d'une saison ou d'une année

particulière, il est admis que les éléments de la série vectorielle  $\varepsilon_m(t)$  ( $t = 1, \dots, n$ ),  $\varepsilon_s(t)$  ( $t = 1, \dots, n$ ) ou  $\varepsilon_y(t)$  ( $t = 1, \dots, n$ ) sont totalement indépendants les uns des autres dans le temps. Toutefois, une fois que les séries ont été examinées en parallèle, il est nécessaire de procéder à une synthèse pour les signaux d'inhomogénéité mensuels  $I\hat{H}_{j,m}(t)$  ( $m = 1, 2, \dots, 12$ ) tels qu'ils ont été estimés. Si les séries de données sont distribuées de façon quasi log-normale (dans le cas des précipitations par exemple), on peut utiliser un modèle multiplicatif susceptible de devenir un modèle additif via une transformation logarithmique.

## 5.2 Méthode de comparaison des séries dans le cas d'un modèle additif

Dans le cas de l'homogénéisation relative, une série de données provenant d'une station candidate fait office de série de référence par rapport à celles des autres stations, ce qui permet d'éliminer le signal commun de changement climatique et de variabilité du climat, qu'on ne connaît pas.

Via la notation de l'équation 1, toutes les séries de données de stations qui sont examinées –  $X_{j,m}(t)$  ( $j = 1, \dots, N$ ) – doivent être considérées à la fois comme des séries à traiter et des séries de référence. On ne présuppose pas par ailleurs que les séries de référence sont homogènes dans la mesure où, en général, on n'en sait rien. La comparaison vise surtout à éliminer le signal de changement climatique  $\mu_m(t)$ .

Les questions afférentes à la comparaison de séries de données sont les suivantes:

- Comparaison par paires;
- Création de séries de référence composites;
- Constitution de séries résultant d'une comparaison par différence (séries de différences);
- Comparaisons multiples de séries.

Ces questions revêtent une grande importance lorsqu'il s'agit de détecter des défauts d'homogénéité et d'y remédier, car une comparaison efficace des séries permet d'accroître la capacité de détection et l'exactitude des corrections. On peut mettre au point des méthodes de comparaison efficaces en examinant la structure de covariance spatiale des séries de données.

Les séries résultant d'une comparaison par différence entre paires sont  $Z_{j,m}(t) = X_{j,m}(t) - X_{i,m}(t)$ . La constitution de ce type de série peut être toutefois formulée aussi de façon plus générale.

En admettant que  $X_{j,m}(t)$  est la série à traiter et que les autres sont les séries de référence, la série de différences qui relève de la série à traiter peut être constituée comme suit :

$$Z_{j,m}(t) = X_{j,m}(t) - \sum_{i \neq j} \lambda_{ji,m} X_{i,m}(t) = IH_{j,m}(t) - \sum_{i \neq j} \lambda_{ji,m} IH_{i,m}(t) + \varepsilon_{Z_{j,m}}(t), \quad (2)$$

où  $\sum_{i \neq j} \lambda_{ji,m} X_{i,m}(t)$  est une série de référence composite avec, comme condition,  $\sum_{i \neq j} \lambda_{ji,m} = 1$  pour

les facteurs de pondération. Il résulte de cette dernière condition que le signal de changement climatique et de variabilité du climat  $\mu_m(t)$ , qu'on ne connaît pas, a été éliminé. Il est donc possible de déceler la rupture d'homogénéité en examinant la série ci-dessus résultant d'une comparaison par différence. Comme celle-ci (2) comporte généralement des défauts d'homogénéité provenant tant de la série à traiter que de la série de référence, il serait peut-être utile de choisir des séries de référence de meilleure qualité et de constituer plusieurs séries de référence composites pour une série à traiter. Des comparaisons multiples ou l'examen de séries multiples résultant d'une comparaison par différence peut aider à déterminer les causes du défaut d'homogénéité détecté pour la série à traiter.

La sélection ou la pondération de la série de référence obéit à trois considérations. Il s'agit en effet :

- 1) De réduire le bruit de la série résultant d'une comparaison par différence;
- 2) De réduire l'influence des problèmes d'homogénéité de la série de référence;
- 3) De veiller à ce que la série de référence ait un signal climatique régional analogue à celui de la série à traiter.

Pour accroître le rapport signal-bruit et renforcer ainsi la capacité de détection, il est nécessaire de réduire la variance du terme qui prend en compte le « bruit de fond »:  $\varepsilon_{Z_{j,m}}(t)$ . Les facteurs de pondération optimaux  $\lambda_{j,m}(t)$  susceptibles de réduire la variance sont déterminés uniquement par les matrices de covariance spatiale  $C_m$  (interpolation optimale ou pondérations de krigeage). Pour limiter l'influence des défauts d'homogénéité des relevés de stations voisines, auxquels sont attribués des facteurs élevés de pondération de krigeage, il peut être souhaitable d'établir une moyenne sur plusieurs stations ou bien d'accorder relativement plus de poids à des stations éloignées. D'autre part, si l'on veut que le signal climatique régional soit le même pour toutes les stations, il est préférable de limiter le nombre de stations considérées lorsque le réseau est peu dense.

Il est prouvé mathématiquement que selon le principe de la probabilité maximale appliqué à la détection et/ou correction commune (dans l'hypothèse d'une distribution gaussienne), les séries résultant d'une comparaison par différence sont examinées indirectement, et les facteurs de pondération de la série de référence sont déterminés par les matrices de covariance spatiale. Cette question est à nouveau abordée dans la section 5.3.1 concernant la sélection de modèle par des méthodes de probabilité pénalisée.

### 5.3 Méthode de détection du point de rupture (point de changement)

Un élément essentiel du processus d'homogénéisation consiste à examiner les séries résultant d'une comparaison par différence (2) afin de détecter les points de rupture et d'attribuer à la série à traiter ceux qui en relèvent.

Le principe de détection du point de rupture est le suivant: postulons que  $Z(t)$  est une série résultant d'une comparaison par différence selon la formule (2), à savoir:

$$Z(t) = IH_z(t) + \varepsilon_z(t) \quad (t = 1, \dots, n), \quad (3)$$

où  $IH_z(t)$  est une inhomogénéité mixte de la série en question  $Z(t)$ , avec des points de rupture provenant de la série à traiter et des séries de référence. En général, le nombre de points de rupture, leur position et leur ampleur ne sont pas connus, et l'on part du principe que  $\varepsilon_z(t)$  est une série gaussienne prenant en compte le « bruit de fond ». Dans le cas d'une homogénéisation menée en parallèle (Remarque N° 1, point 1.2),  $\varepsilon_z(t)$  est une série gaussienne de type bruit blanc (sans corrélation) vu que ses éléments sont supposés indépendants les uns des autres dans le temps.

#### Remarque N° 2

La détection des valeurs aberrantes est la principale méthode de contrôle de la qualité des données mensuelles dans ce contexte. Si une valeur aberrante n'est pas éliminée, elle peut donner lieu à deux points d'inflexion voisins dont l'ampleur est la même en valeur absolue mais de signes opposés.

Pour en revenir aux méthodes de détection, les deux principales sont la classification hiérarchique (détection par étapes) et la détection des points de rupture multiples. La première consiste à détecter l'un après l'autre chaque point de rupture et la seconde vise à estimer tous les points de rupture d'une série à traiter.

Il existe deux méthodes traditionnelles de détection des points de rupture, qui relèvent des statistiques mathématiques, à savoir l'estimation de vraisemblance maximale et les tests d'hypothèses.

#### 5.3.1 **Détection des points de rupture fondée sur l'estimation de vraisemblance maximale**

Les méthodes de probabilité pénalisée reposent sur la sélection de modèle (segmentation): une estimation de vraisemblance maximale conjointe est donnée pour les points de rupture en présupposant une distribution gaussienne de la série résultant d'une comparaison par différence et moyennant un terme pénalisant, dont l'utilisation se justifie par le fait qu'on ne connaît pas le nombre de points d'inflexion. Les méthodes peuvent différer selon les critères ou les termes pénalisants, par exemple le critère d'information d'Akaike (AIC), le critère d'information bayésien (BIC) ou Schwarz ou encore le critère de Caussinus et Lyazrhi. Les termes pénalisants dépendent de la probabilité *a priori* d'une rupture à chaque fois.

Théoriquement, cette méthode pourrait être appliquée aussi à la détection conjointe des points de rupture de toutes les séries examinées. Toutefois, si l'on postule une distribution gaussienne, la fonction de probabilité conjointe dépend de l'inverse de la matrice de covariance spatiale, ce qui peut donner lieu à des problèmes techniques complexes pour les plus grands réseaux. Un problème que peut poser un modèle statistique non trivial est que les signaux climatiques peuvent être différents dans les plus grands réseaux.

### 5.3.2 **Détection des points de rupture fondée sur les tests d'hypothèses**

Le test d'hypothèse est une autre méthode de détection des points de rupture dans une série résultant d'une comparaison par différence. L'hypothèse nulle signifie que la série testée est homogène et qu'il s'agit d'une série gaussienne de type bruit blanc. Cette méthode suppose également une distribution gaussienne, c'est pourquoi les statistiques de test sont des statistiques de type t en général. Si aucune série de référence n'est utilisée (homogénéisation absolue), les statistiques de type F sont dérivées en raison de la régression de la tendance temporelle qui n'est pas connue. L'importance et la capacité de cette méthode peuvent être définies en fonction des probabilités de deux types d'erreurs: les erreurs de type 1, qui consistent à détecter de faux points de rupture, et les erreurs de type 2, qui consistent à ne pas détecter les points de rupture qui existent réellement. Il convient de trouver un compromis entre ces deux types d'erreurs. Les statistiques de test peuvent être comparés à la valeur critique qui dépend du niveau de signification donné. Dans le cas de statistiques de test multiples, les valeurs critiques peuvent être calculées à l'aide des méthodes de Monte-Carlo.

La plupart de ces méthodes sont des méthodes de détection par étapes, de chaque point de rupture pris un par un. Toutefois, des méthodes de détection de tous les points de rupture d'une série peuvent également être élaborées. Cette procédure obéit au principe selon lequel l'homogénéité peut être acceptée entre des points de rupture voisins détectés, mais ne peut l'être entre des points de rupture qui ne sont pas voisins. En outre, des intervalles de confiance peuvent également être donnés pour les points de rupture, qui rendent possible l'utilisation automatique des métadonnées.

Ces méthodes ont l'avantage de permettre d'évaluer et de valider les résultats par comparaison avec les statistiques de test avant et après l'homogénéisation

### 5.3.3 **Attribution des points de rupture détectés pour les séries à traiter**

Au cours de la procédure de détection des points de rupture, les séries résultant d'une comparaison par différence sont examinées avec des défauts d'homogénéité mixtes (équation 2) Il existe deux principaux moyens de résoudre ce problème:

1. Si une seule série de différences est examinée pour une série à traiter, toutes les ruptures détectées sont attribuées à la série à traiter. Dans ce cas, les défauts d'homogénéité qui peuvent exister dans la série de référence constituent un grave problème. Il convient par conséquent de choisir davantage de séries de référence de meilleure qualité afin de constituer une série de référence composite, de préférence sans point de rupture (il est recommandé d'utiliser au moins quatre stations de référence);
2. Si plusieurs séries de différences sont examinées pour une série à traiter, la présence de défauts d'homogénéité mixtes est moins problématique, mais l'attribution des points de rupture à la série à traiter n'est pas une tâche aisée. Il est donc nécessaire de procéder à une synthèse et le logiciel d'homogénéisation a pour principale fonction d'appliquer des procédures automatiques pour l'attribution des points de rupture.

## 5.4 **Méthodes applicables pour l'ajustement des séries**

Outre la détection, l'ajustement des séries constitue une autre activité essentielle. On peut calculer les facteurs d'ajustement en procédant à l'examen des séries de différences afin d'estimer les écarts aux points de rupture détectés. On a recours en général dans ce cas à une estimation ponctuelle.

Certaines méthodes utilisent la technique standard des moindres carrés une fois les ruptures détectées pour une estimation conjointe des modifications de toutes les séries examinées. La technique d'estimation des moindres carrés généralisés basée sur la matrice de covariance spatiale pourrait être la technique la plus efficace dans ce cas, et elle serait équivalente à l'estimation de vraisemblance maximale pour les modifications dans le cas de la distribution gaussienne. Il convient cependant de résoudre un problème au niveau du modèle statistique non trivial à savoir que les signaux climatiques peuvent être différents dans les plus grands réseaux.

Une autre méthode consiste à calculer les facteurs d'ajustement sur la base de certains intervalles de confiance donnés pour les modifications aux points de rupture détectés. Ces intervalles de confiance permettent également l'utilisation automatique des métadonnées.

#### Remarque N° 3

Compléter les données manquantes ou combler les lacunes constitue essentiellement un problème d'interpolation. Il est vivement recommandé d'utiliser les covariances spatiales pour calculer les facteurs de pondération des prédicteurs afin de diminuer les erreurs d'interpolation.

### 5.5 Options applicables pour l'évaluation et la validation des méthodes

Pour bien comprendre les méthodes disponibles, il est indispensable de procéder à une évaluation théorique de leur base mathématique.

Une autre option consiste à effectuer une étude de comparaison et de validation en aveugle des méthodes d'homogénéisation. Dans ce cas, les méthodes sont testées sur un jeu de données de référence réalistes. Ce jeu de référence contient des données simulées dans lesquelles on a inséré un défaut d'homogénéité. Tester les méthodes à partir d'un jeu de données de référence constitué semble être une procédure de validation objective, mais ce type d'évaluation a ses limites.

L'interprétation des résultats dépend de différents facteurs, tels que:

- Les méthodes testées (qualité, manuelles ou automatiques);
- L'évaluation du jeu de données de référence (qualité, adéquation);
- Les opérateurs (qualifiés ou non);
- La méthode d'évaluation (statistiques de validation).

La création d'un jeu de données de référence adéquat et l'élaboration de statistiques de validation appropriées sont des activités essentielles. Elles nécessitent une solide formation théorique en mathématiques, par exemple, pour comprendre quelles sont les caractéristiques statistiques d'un point de référence qui sont importantes, et la modélisation doit répondre à des critères objectifs, nécessitant éventuellement une étude plus détaillée pour déterminer quelle marge est applicable dans les réseaux réels (Lindau et Venema, 2019).

L'un des avantages des méthodes d'homogénéisation automatiques est qu'elles se prêtent à une validation objective pour deux raisons : a) la validation objective nécessite des jeux de données relativement importants pour être performante et les méthodes d'homogénéisation automatiques permettent d'homogénéiser plus facilement ces grands jeux de données; b) dans le cas des méthodes manuelles, non seulement la méthode mais aussi l'opérateur sont évalués, ce qui peut fausser les résultats dans un sens ou dans l'autre.



## GLOSSAIRE

**Abri.** Boîtier destiné à protéger les instruments météorologiques afin d'enregistrer correctement les conditions atmosphériques conformément aux normes de l'OMM. Il existe de nombreux types d'abris différents.

**AIC.** Critère d'information d'Akaike, une fonction de pénalisation.

**Analyse de la variance (ANOVA).** Méthode de correction conjointe qui consiste à calculer simultanément toutes les corrections d'un réseau en supposant que toutes ont le même signal climatique régional et que les ruptures correspondent à une variation d'un échelon. Cette méthode minimise le bruit en utilisant une technique des moindres carrés, de sorte que les équations sont les mêmes que celles du test statistique pour les différences de moyennes.

**BIC.** Critère d'information bayésien, une fonction de pénalisation.

**Correction.** Ajustement des données d'observation brutes visant à les rendre plus homogènes.

**Évaluation comparative.** Test de performance des méthodes d'homogénéisation utilisant des données ouvertes réalistes, ce qui implique en général plusieurs méthodes, contributions ou opérateurs. S'inscrivant dans le cadre d'un effort collectif, l'évaluation comparative est plus qu'un simple test de validation d'une méthode.

**HadCRUT.** Jeu de données sur la température mondiale compilé par le Centre Hadley du Service météorologique du Royaume-Uni et l'Unité de recherche sur le climat (CRU) de l'Université d'East Anglia.

**Homogénéisation annuelle.** Homogénéisation utilisant des données annuelles (moyennes ou sommes). Elle peut également inclure le poids du cycle saisonnier.

**Homogénéisation mensuelle.** Homogénéisation utilisant des données mensuelles (moyennes ou sommes). Cela peut se faire avec 12 séries chronologiques mensuelles en parallèle (par défaut) ou avec tous les mois considérés comme une seule et longue série chronologique. Souvent combinée avec une homogénéisation annuelle.

**Homogénéisation quotidienne.** Homogénéisation utilisant des données journalières (moyennes ou sommes). Il s'agit en général de considérer toutes les données journalières comme une seule et longue série chronologique.

**Homogénéisation saisonnière.** Homogénéisation utilisant des données saisonnières (moyennes ou sommes). Cela peut se faire en mettant en parallèle quatre séries chronologiques saisonnières (par défaut) ou en considérant toutes les saisons comme une seule et longue série chronologique. Souvent combinée avec une homogénéisation annuelle.

**iCOADS.** Ensemble international intégré de données sur l'océan et l'atmosphère.

**Méthode des points de rupture multiples.** Méthode d'homogénéisation qui détecte ou corrige de multiples défauts d'homogénéité dans une station/série chronologique au cours d'une même étape.

**Méthodes conjointes.** Méthodes d'homogénéisation qui détectent ou corrigent conjointement tous les défauts d'homogénéité dans plusieurs stations au cours d'une même étape.

**PMT.** *Penalized Maximal t Test* (Test *t* maximal pénalisé).

**PRODIGE.** Méthode d'homogénéisation. Méthode des points de rupture multiples basée sur le critère de Caussinus-Lyazrhi.

**Série chronologique de référence.** Série chronologique indépendante utilisée pour évaluer l'homogénéité d'une station candidate, généralement issue d'une station voisine ou dérivée de plusieurs stations voisines.

**Sous-période homogène (HSP).** Un segment entre des points de changement adjacents pour lequel les données pourraient être considérées comme étant homogènes (entre les inhomogénéités).

**Station candidate/de base.** Station dont les séries de données doivent être homogénéisées.

**UTC.** Temps universel coordonné.

---

## RÉFÉRENCES

- Alexandersson, H., 1986: A homogeneity test applied to precipitation data. *Journal of Climatology*, 6, pp. 661-675, <https://doi.org/10.1002/joc.3370060607>.
- Azarin-Molina, C. *et al.*, 2014: Homogenization and assessment of observed near-surface wind speed trends over Spain and Portugal, 1961-2011. *Journal of Climate*, 27, pp. 3692-3712, <https://doi.org/10.1175/JCLI-D-13-00652.1>.
- Azarin-Molina, C. *et al.*, 2019: An approach to homogenize daily peak wind gusts: An application to the Australian series. *International Journal of Climatology*, 18 p., <https://doi.org/10.1002/JOC.5949>.
- Begert M., T. Schlegel et W. Kirchhofer, 2005: Homogeneous temperature and precipitation series of Switzerland from 1864 to 2000. *International Journal of Climatology*, 25, pp. 65-80, <https://dx.doi.org/10.1002/joc.1118>.
- Böhm, R. *et al.*, 2010: The early instrumental warm-bias: a solution for long central European temperature series 1760-2007. *Climatic Change*, 101, pp. 41-67, <https://link.springer.com/article/10.1007/s10584-009-9649-4>
- Brandsma, T., J. Van der Meulen et V.K.C. Venema, 2019: Protocol Measurement Infrastructure Changes (PMIC). Publié prochainement en tant que rapport de l'OMM.
- Brogniez, H. *et al.*, 2016: A review of sources of systematic errors and uncertainties in observations and simulations at 183 GHz. *Atmospheric Measurement Techniques*, 9, pp. 2207-2221, <https://doi.org/10.5194/amt-9-2207-2016>.
- Brückner, E., 1890: *Klimaschwankungen seit 1700 nebst Bemerkungen über Klimaschwankungen der Diluvialzeit*. E.D. Hölzel, Wien et Olnütz.
- Brugnara Y. *et al.*, 2012: High-resolution analysis of daily precipitation trends in the central Alps over the last century. *International Journal of Climatology*, 32, pp. 1406-1422, <https://doi.org/10.1002/joc.2363>.
- Brunet, M. *et al.*, 2006: The development of a new dataset of Spanish daily adjusted temperature series (SDATS) (1850-2003). *International Journal of Climatology*, 26, pp. 1777-1802, <https://doi.org/10.1002/joc.1338>.
- Brunet, M. *et al.*, 2011: The minimisation of the screen bias from ancient Western Mediterranean air temperature records: an exploratory statistical analysis. *International Journal of Climatology*, 31, pp. 1879-1895, <https://doi.org/10.1002/joc.2192>.
- Brunetti M. *et al.*, 2006: Temperature and precipitation variability in Italy in the last two centuries from homogenized instrumental time series. *International Journal of Climatology*, 26, pp. 345-381, <https://doi.org/10.1002/joc.1251>.
- Buisan, S.T., C. Azarin-Molina et Y. Jimenez, 2015: Impact of two different sized Stevenson screens on air temperature measurements. *International Journal of Climatology*, 35(14), pp. 4408-4416, <https://doi.org/10.1002/joc.4287>.
- Caussinus, H. et O. Mestre, 2004: Detection and correction of artificial shifts in climate series. *Journal of the Royal Statistical Society, Series C (Applied Statistics)*, 53, pp. 405-425, <https://doi.org/10.1111/j.1467-9876.2004.05155.x>.
- Conrad, V., 1925: Homogenitätsbestimmung meteorologischer Beobachtungsreihen. *Meteorologische Zeitschrift*, pp. 482-485.
- Conrad V. et O. Schreier, 1927: Die Anwendung des Abbe'schen Kriteriums auf physikalische Beobachtungsreihen. *Beiträge zur Geophysik de Gerland*, XVII, 372.
- Conrad V., 1944: *Methods in Climatology*. Harvard University Press, 228 p.
- Conrad, V. et C. Pollak, 1950: *Methods in Climatology*. Harvard University Press, Cambridge, MA, 459 p.
- Costa, A.C. et A. Soares, 2009: Homogenization of climate data: Review and new perspectives using Geostatistics. *Mathematical Geosciences*, 41, pp. 291-305, <https://doi.org/10.1007/s11004-008-9203-3>.
- Cowtan, K., R. Rohde et Z. Hausfather, 2018: Evaluating biases in sea surface temperature records using coastal weather stations. *Quarterly Journal of the Royal Meteorological Society*, <https://doi.org/10.1002/qj.3235>.
- Cook, B.I. *et al.*, 2014: Irrigation as an historical climate forcing. *Climate Dynamics*, 44: 1715, <https://doi.org/10.1007/s00382-014-2204-7>.
- Craddock, J.M., 1979: Methods of comparing annual rainfall records for climatic purposes. *Weather*, 34, pp. 332-346, <https://doi.org/10.1002/j.1477-8696.1979.tb03465.x>.
- Dai, A. *et al.*, 2011: A new approach to homogenize daily radiosonde humidity data. *Journal of Climate*, 24, 965991, <https://doi.org/10.1175/2010JCLI3816.1>.

- Degaetano, A.T., 2000: A serially complete simulated observation time metadata file for U.S. daily Historical Climatology Network stations. *Bulletin of the American Meteorological Society*, 81: 1, pp. 49–68, [https://doi.org/10.1175/1520-0477\(2000\)081<0049:ASCSOT>2.3.CO;2](https://doi.org/10.1175/1520-0477(2000)081<0049:ASCSOT>2.3.CO;2).
- Della-Marta, P.M. et H. Wanner, 2006: A method of homogenizing the extremes and mean of daily temperature measurements. *Journal of Climate*, 19, pp. 4179–4197. <https://doi.org/10.1175/JCLI3855.1>.
- Dienst, M. *et al.*, 2017: Removing the relocation bias from the 155-year Haparanda temperature record in Northern Europe. *International Journal of Climatology*, 37: pp. 4015–4026, <https://doi.org/10.1002/joc.4981>.
- Dienst, M. *et al.*, 2019: Detection and elimination of UHI effects in long temperature records from villages – A case study from Tivissa, Spain. *Urban Climate*, 27, pp. 372–383, <https://doi.org/10.1016/j.uclim.2018.12.012>.
- Domonkos, P., 2011: Efficiency evaluation for detecting inhomogeneities by objective homogenization methods. *Theoretical and Applied Climatology*, 105, pp. 455–467. <https://doi.org/10.1007/s00704-011-0399-7>.
- Domonkos, P. et J. Coll, 2017: Homogenisation of temperature and precipitation time series with ACMANT3: Method description and efficiency tests. *International Journal of Climatology*, 37, pp. 1910–1921, <https://doi.org/10.1002/joc.4822>.
- Domonkos, P., V. Venema et O. Mestre, 2013: Efficiencies of homogenisation methods: our present knowledge and its limitation. In *Seventh Seminar for Homogenization and Quality Control in Climatological Databases Jointly Organized with the COST ES0601 (HOME) Action MC Meeting*, Budapest, Hungary, 24–27 October 2011. *Climate data and monitoring* (WCDMP-No. 78), Genève, OMM, pp. 11–24.
- Dunn, R. et P. Thorne, 2017: Towards an integrated set of surface meteorological observations for climate science and applications. *Compte rendu de la dix-neuvième Assemblée générale de l'Union européenne des géosciences*, Vienne, Autriche.
- Dunn, R.J.H. *et al.*, 2012: HadISD: a quality-controlled global synoptic report database for selected variables at long-term stations from 1973–2011. *Climate of the Past*, 8, pp. 1649–1679, <https://doi.org/10.5194/cp-8-1649-2012>.
- Gubler, S. *et al.*, 2017: The influence of station density on climate data homogenization. *International Journal of Climatology*, 37, pp. 4670–4683, <https://doi.org/10.1002/joc.5114>.
- Guijarro, J.A., 2013: Temperature trends. In *Adverse weather in Spain* (dir. publ.: C. García-Legaz C et F. Valero). Madrid, AMV Ediciones, pp. 297–306.
- Guijarro, J., 2018: Climatol, Version 3.1.2, <https://CRAN.R-project.org/package=climatol> et <http://www.climatol.eu/>.
- Haimberger, L., 2007: Homogenization of radiosonde temperature time series using innovation statistics. *Journal of Climate*, 20(7): pp. 1377–1403, <https://doi.org/10.1175/JCLI4050.1>.
- Haimberger, L., C. Tavalato et S. Sperka, 2012: Homogenization of the global radiosonde dataset through combined comparison with reanalysis background series and neighboring stations. *Journal of Climate*, 25, pp. 8108–8131, <https://doi.org/10.1175/JCLI-D-11-00668.1>.
- Hann, J., 1880: *Untersuchungen über die Regenverhältnisse von Österreich-Ungarn. II. Veränderlichkeit der Monats- und Jahresmengen*. S.-B. Akad. Wiss. Wien.
- Heidke, P., 1923: Quantitative Begriffsbestimmung homogener Temperatur- und Niederschlagsreihen. *Meteorologische Zeitschrift*, pp. 114–115.
- Helmert, F.R., 1907: *Die Ausgleichrechnung nach der Methode der kleinsten Quadrate*. 2. Auflage, Teubner Verlag.
- Huang, B. *et al.*, 2015: Extended Reconstructed Sea Surface Temperature Version 4 (ERSST.v4). Part I: Upgrades and Intercomparisons. *Journal of Climate*, 28, pp. 911–930, <https://doi.org/10.1175/JCLI-D-14-00006.1>.
- Jones, P.D. *et al.*, 2012: Hemispheric and large-scale land-surface air temperature variations: An extensive revision and an update to 2010. *Journal of Geophysical Research*, 117: D05127, <https://doi.org/10.1029/2011JD017139>.
- Jovanovic, B. *et al.*, 2017: Homogenized monthly upper-air temperature dataset for Australia. *International Journal of Climatology*, 37, pp. 3209–3222, <https://doi.org/10.1002/joc.4909>.
- Karl, T.R. *et al.*, 1986: A model to estimate the time of observation bias associated with monthly mean maximum, minimum and mean temperatures for the United States, *Journal of Applied Meteorology and Climatology*, 25:2, pp. 145–160, [https://doi.org/10.1175/1520-0450\(1986\)025<0145:AMTETT>2.0.CO;2](https://doi.org/10.1175/1520-0450(1986)025<0145:AMTETT>2.0.CO;2).

- Kennedy, J.J. *et al.*, 2011: Reassessing biases and other uncertainties in sea surface temperature observations measured in situ since 1850: 1. Measurement and sampling uncertainties. *Journal of Geophysical Research – Atmospheres*, 116:D14103, <https://doi.org/10.1029/2010JD015218>.
- Kent, E. *et al.*, 2016: A call for new approaches to quantifying biases in observations of sea-surface temperature. *Bulletin of the American Meteorological Society*, 98, pp. 1601–1616, <https://doi.org/10.1175/BAMS-D-15-00251.1>.
- k.k. Hof- und Staatsdruckerei, 1873: *Bericht über die Verhandlungen des internationalen Meteorologen-Congresses zu Wien*, 2–10. septembre 1873, Protokolle und Beilagen.
- k.k. Zentralanstalt für Meteorologie und Geodynamik, 1906: *Bericht über die internationale meteorologische Direktorenkonferenz in Innsbruck*, septembre 1905. Anhang zum Jahrbuch 1905. k.k. Hof- und Staatsdruckerei.
- Kohler M.A., 1949: On the use of double-mass analysis for testing the consistency of records and for making required adjustments. *Bulletin of the American Meteorological Society*, 30, pp. 188–189, <https://doi.org/10.1175/1520-0477-30.5.188>.
- Kuglitsch, F.G., 2012: Break detection of annual Swiss temperature series. *Journal of Geophysical Research – Atmospheres*, 117:D13105, <https://doi.org/10.1029/2012JD017729>.
- Kreil, K., 1854a: Mehrjährige Beobachtungen in Wien vom Jahre 1775 bis 1850. *Jahrbücher der k.k. Central-Anstalt für Meteorologie und Erdmagnetismus*. I. Band – Jg 1848 und 1849, pp. 35–74.
- Kreil, K., 1854b: Mehrjährige Beobachtungen in Mailand vom Jahre 1763 bis 1850. *Jahrbücher der k.k. Central-Anstalt für Meteorologie und Erdmagnetismus*. I. Band – Jg 1848 und 1849, pp. 75–114.
- Leeper, R.D., J. Rennie et M.A. Palecki, 2015: Observational perspectives from U.S. Climate Reference Network (USCRN) and Cooperative Observer Program (COOP) network: Temperature and precipitation comparison. *Journal of Atmospheric and Oceanic Technology*, 32, pp. 703–721, <https://doi.org/10.1175/JTECH-D-14-00172.1>.
- Li, Y., R. Lund et A. Hewaarachchi, 2017: Multiple changepoint detection with partial information on changepoint times. *ArXiv:1511.07238* (manuscript), <https://arxiv.org/abs/1511.07238>.
- Lindau, R. et V.K.C. Venema, 2013: On the multiple breakpoint problem and the number of significant breaks in homogenization of climate records. *Idojaras*, 117(1):1, pp. 1–34.
- Lindau, R. et V. Venema, 2016: The uncertainty of break positions detected by homogenization algorithms in climate records. *International Journal of Climatology*, 36(2), pp. 576–589, <https://doi.org/10.1002/joc.4366>.
- Lindau, R. et V.K.C. Venema, 2018a: The joint influence of break and noise variance on the break detection capability in time series homogenization. *Advances in Statistical Climatology, Meteorology and Oceanography*, 4, pp. 1–18. <https://doi.org/10.5194/ascmo-4-1-2018>.
- Lindau, R. et V.K.C. Venema, 2018b: On the reduction of trend errors by the ANOVA joint correction scheme used in homogenization of climate station records. *International Journal of Climatology*, 38(14), pp.5255–5271, <https://doi.org/10.1002/joc.5728>.
- Lindau, R. et V.K.C. Venema, 2019: A new method to study inhomogeneities in climate records: Brownian motion or random deviations? *International Journal of Climatology*, 39(12), <https://doi.org/10.1002/joc.6105>.
- Lu, Q.Q., R. Lund et T.C.M. Lee, 2010: An MDL approach to the climate segmentation problem. *The Annals of Applied Statistics*, 4(1), 299–319, <https://doi.org/10.1214/09-AOAS289>.
- Lund, R. et J. Reeves, 2002: Detection of undocumented changepoints: A revision of the two-phase regression model. *Journal of Climate*, 15, pp. 2547–2554, [https://doi.org/10.1175/1520-0442\(2002\)015%3C2547:DOUCAR%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C2547:DOUCAR%3E2.0.CO;2).
- Mekis, É. et L.A. Vincent, 2011: An overview of the second generation adjusted daily precipitation dataset for trend analysis in Canada. *Atmosphere-Ocean*, 49:2, pp. 163–177, <https://doi.org/10.1080/07055900.2011.583910>.
- Menne, M.J., C.N. Williams Jr. et M.A. Palecki, 2010: On the reliability of the U.S. surface temperature record. *Journal of Geophysical Research – Atmospheres*, 115:D11, <https://doi.org/10.1029/2009JD013094>.
- Menne, M.J. et C.N. Williams Jr., 2009: Homogenization of temperature series via pairwise comparisons. *Journal of Climate*, 22:7, pp. 1700–1717, <https://doi.org/10.1175/2008JCLI2263.1>.
- Menne, M.J., C.N. Williams Jr. et R.S. Vose, 2009: The U.S. Historical Climatology Network monthly temperature data, version 2. *Bulletin of the American Meteorological Society*, 90, pp. 993–1007, <https://doi.org/10.1175/2008BAMS2613.1>.
- Mestre O., 1999: Step-by-step procedures for choosing a model with change-points. In *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, (9–13 November 1998, Budapest, Hungary) (WCDMP-No.41, WMO/TD-No. 962). Genève, OMM, pp. 15–26.

- Mestre, O. *et al.*, 2011: SPLIDHOM: A method for homogenization of daily temperature observations. *Journal of Applied Meteorology and Climatology*, 50, pp. 2343–2358, <https://doi.org/10.1175/2011JAMC2641.1>.
- Mestre, O. *et al.*, 2013: HOMER: A homogenization software – Methods and applications. *Idojaras*, 117, 47–67.
- Minola, L., C. Azorin-Molina et D. Chen, 2016: Homogenization and assessment of observed near-surface wind speed trends across Sweden, 1956–2013. *Journal of Climate*, 29, pp. 7397–7415, <https://doi.org/10.1175/JCLI-D-15-0636.1>.
- Organisation météorologique mondiale (OMM), 1966: *Climatic Change: Report of a Working Group of the Commission for Climatology* (WMO–No. 195). Genève.
- , 1999: *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, 9–13 November 1998 (WMO/TD–No. 962; WCDMP–No. 41). Genève.
- , 2004: *Fourth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, 6–10 October 2003 (WMO/TD–No. 1236, WCDMP–No. 56). Genève.
- , 2006: *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, 29 May–2 June 2006 (WMO/TD–No. 1493, WCDMP–No. 71). Genève.
- , 2010: *Proceedings of the Sixth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, 26–30 May 2008 (WMO/TD–No. 1576, WCDMP–No. 76). Genève.
- , 2011: *Seventh Seminar for Homogenization and Quality Control in Climatological Databases jointly organized with the COST E2S0601 (HOME) Action MC Meeting*, Budapest, Hungary, 24–27 October 2011 (dir. publ.: M. Lakatos, T. Szentimrey et E. Vincze) (WCDMP–No. 78). Genève.
- , 2014: *Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques in Climatology and Meteorology*, Budapest, Hungary, 12–16 May 2014 (dir. publ.: M. Lakatos, T. Szentimrey et A. Marton) (WCDMP–No. 84). Genève.
- , 2019: *Norme relative aux métadonnées du WIGOS* (OMM–N° 1192). Genève.
- , 2018a: *Guide des instruments et des méthodes d'observation* (OMM–N° 8). Genève.
- , 2018b: *Guide des pratiques climatologiques* (OMM–N° 100). Genève.
- Parker, D.E., 1994: Effects of changing exposure of thermometers at land stations. *International Journal Climatology*, 14, pp. 1–31, <https://doi.org/10.1002/joc.3370140102>.
- Petrovic, P., 2004: Detecting of inhomogeneities in time series using Real Precision Method. In: *Fourth Seminar for Homogenization and Quality Control in Climatological Databases* (WCDMP–No. 56, WMO/TD–No. 1236). Genève, OMM.
- Quayle, R.G. *et al.*, 1991: Effects of recent thermometer changes in the cooperative station network. *Bulletin of the American Meteorological Society*, 72, pp. 1718–1723. [https://doi.org/10.1175/1520-0477\(1991\)072%3C1718:EORTCI%3E2.0.CO;2](https://doi.org/10.1175/1520-0477(1991)072%3C1718:EORTCI%3E2.0.CO;2).
- Rennie, J.J., 2014: The international surface temperature initiative global land surface databank: Monthly temperature data release description and methods. *Geoscience Data Journal*, 1:2, pp. 75–102, <https://doi.org/10.1002/gdj3.8>.
- Ribeiro, S. *et al.*, 2017: GSIMCLI: A geostatistical procedure for the homogenization of climatic time series. *International Journal of Climatology*, 37:8, pp. 3452–3467, <https://doi.org/10.1002/joc.4929>.
- Rohde, R. *et al.*, 2013: Berkeley Earth Temperature Averaging Process. *Geoinformatics & Geostatistics: An Overview*, 1:2, <https://doi.org/10.4172/2327-4581.1000103>.
- Schröder, M. *et al.*, 2016: The GEWEX water vapor assessment: Results from intercomparison, trend, and homogeneity analysis of total column water vapor. *Journal of Applied Meteorology and Climatology*, 55, pp. 1633–1649, <https://doi.org/10.1175/JAMC-D-15-0304.1>.
- Service météorologique hongrois (OMSZ), 1996: *Proceedings of the First Seminar for Homogenization of Surface Climatological Data*, Budapest, Hongrie, 6–12 octobre 1996, 44 p.
- Service météorologique hongrois (OMSZ), 2001: *Third Seminar for Homogenization and Quality Control in Climatological Databases*. Budapest.
- Service météorologique hongrois (OMSZ), 2017: *9th Seminar for Homogenization and Quality Control in Climatological Databases and 4th Conference on Spatial Interpolation Techniques in Climatology and Meteorology*, Budapest, Hungary, 3–7 April 2017 (dir. publ.: T. Szentimrey, L. Hoffmann et M. Lakatos). Budapest, <https://doi.org/10.21404/9.SemHQC4.ConfSI.2017>.
- Štěpánek, P., P. Zahradníček et P. Skalák 2009: Data quality control and homogenization of the air temperature and precipitation series in the Czech Republic in the period 1961–2007. *Advances in Science and Research*, 3, pp. 23–26, <https://doi.org/10.5194/asr-3-23-2009>.
- Szentimrey, T., 1999: Multiple Analysis of Series for Homogenization (MASH). In *Proceedings of the Second Seminar for Homogenization of Surface Climatological Data*, Budapest, Hungary, 9–13 November 1998 (WCDMP–No. 41, WMO/TD–No. 962). Genève, OMM, pp. 27–46.

- Szentimrey, T., 2008: Development of MASH homogenization procedure for daily data. In *Proceedings of the Fifth Seminar for Homogenization and Quality Control in Climatological Databases*, Budapest, Hungary, 29 May–2 June 2006 (WCDMP-No. 71, WMO/TD-No. 1493). Genève, OMM, pp. 123–130.
- Szentimrey, T., 2014: *Manual of homogenization software MASHv3.03*. Hungarian Meteorological Service, 71 p.
- Szentimrey, T., 2018: New version MASHv4.01 for joint homogenization of mean and standard deviation (EMS Annual Meeting: European Conference for Applied Meteorology and Climatology 2018, Budapest, Hungary, 3–7 September 2018). *EMS Annual Meeting Abstracts*, Vol. 15, EMS2018-331.
- Thorne, P.W. et al., 2011: Guiding the creation of a comprehensive surface temperature resource for twenty-first-century climate science. *Bulletin of the American Meteorological Society*, 92, ES40–ES47, <https://doi.org/10.1175/2011BAMS3124.1>.
- Toreti A. et al., 2010: A novel method for the homogenization of daily temperature series and its relevance for climate change analysis. *Journal of Climate*, 23, pp. 5325–5331, <https://doi.org/10.1175/2010JCLI3499.1>.
- Toreti, A. et al., 2012: A novel approach for the detection of inhomogeneities affecting climate time series. *Journal of Applied Meteorology and Climatology*, 51, 317–326, <https://doi.org/10.1175/JAMC-D-10-05033.1>.
- Trewin, B.C. et A.C.F. Trevitt, 1996: The development of composite temperature records. *International Journal of Climatology*, 16, pp. 1227–1242, [https://doi.org/10.1002/\(SICI\)1097-0088\(199611\)16:11%3C1227::AID-JOC82%3E3.0.CO;2-P](https://doi.org/10.1002/(SICI)1097-0088(199611)16:11%3C1227::AID-JOC82%3E3.0.CO;2-P).
- Trewin, B.C., 2012: *Techniques involved in developing the Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) dataset*. CAWCR Technical Report 49. Melbourne, Centre for Australian Weather and Climate Research, [http://cawcr.gov.au/technical-reports/CTR\\_049.pdf](http://cawcr.gov.au/technical-reports/CTR_049.pdf).
- Trewin, B., 2013. A daily homogenized temperature dataset for Australia. *International Journal of Climatology*, 33:6, pp. 1510–1529, <https://doi.org/10.1002/joc.3530>.
- Trewin, B.C., 2018: *The Australian Climate Observations Reference Network – Surface Air Temperature (ACORN-SAT) version 2*. Bureau Research Report – BRR032. Melbourne, Service météorologique australien.
- Venema, V.K.C. et al., 2012: Benchmarking homogenization algorithms for monthly data. *Climate of the Past*, 8, pp. 89–115, <https://doi.org/10.5194/cp-8-89-2012>.
- Vincent, L.A., 1998: A technique for the identification of inhomogeneities in Canadian temperature series. *Journal of Climate*, 11, pp. 1094–1104, [https://doi.org/10.1175/1520-0442\(1998\)011<1094:ATFTIO>2.0.CO;2](https://doi.org/10.1175/1520-0442(1998)011<1094:ATFTIO>2.0.CO;2).
- Vincent, L.A. et al., 2002: Homogenization of daily temperatures over Canada. *Journal of Climate*, 15, 1322–1334, [https://doi.org/10.1175/1520-0442\(2002\)015%3C1322:HODTOC%3E2.0.CO;2](https://doi.org/10.1175/1520-0442(2002)015%3C1322:HODTOC%3E2.0.CO;2).
- Vincent, L.A. et al., 2009: Bias in minimum temperature introduced by a redefinition of the climatological day at the Canadian Synoptic stations. *Journal of Applied Meteorology and Climatology*, 48, pp. 2160–2168. <https://doi.org/10.1175/2009JAMC2191.1>.
- Vincent, L.A. et al., 2012: A second generation of homogenized Canadian monthly surface air temperature for climate trend analysis. *Journal of Geophysical Research – Atmospheres*, 117:D18110, <https://doi.org/10.1029/2012JD017859>.
- Vose, R.S. et al., 2003: An evaluation of the time of observation bias adjustment in the U.S. Historical Climatology Network. *Geophysical Research Letters*, 30:20, 2046, <https://doi.org/10.1029/2003GL018111>.
- Wan, H., X.L. Wang et V.R. Swail, 2007: A quality assurance system for Canadian hourly pressure data. *Journal of Applied Meteorology and Climatology*, 46, 1804–1817, <https://doi.org/10.1175/2007JAMC1484.1>.
- Wan, H., X.L. Wang et V.R. Swail, 2010: Homogenization and trend analysis of Canadian near-surface wind speeds. *Journal of Climate*, 23, pp. 1209–1225, <https://doi.org/10.1175/2009JCLI3200.1>.
- Wang, X.L., 2003: Comments on “Detection of undocumented change-points: A revision of the two-phase regression model”. *Journal of Climate*, 16, pp. 3383–3385, [10.1175/1520-0442\(2003\)016<3383:CODOUC>2.0.CO;2](https://doi.org/10.1175/1520-0442(2003)016<3383:CODOUC>2.0.CO;2).
- Wang, X.L., Q.H. Wen et Y. Wu, 2007: Penalized maximal *t* test for detecting undocumented mean change in climate data series. *Journal of Applied Meteorology and Climatology*, 46, pp. 916–931, <https://doi.org/10.1175/JAM2504.1>.
- Wang, X.L., 2008a: Accounting for autocorrelation in detecting mean shifts in climate data series using the penalized maximal *t* or *F* test. *Journal of Applied Meteorology and Climatology*, 47, pp. 2423–2444, <https://doi.org/10.1175/2008JAMC1741.1>.

- Wang, X.L., 2008b: Penalized maximal  $F$  test for detecting undocumented mean shift without trend change. *Journal of Atmospheric and Oceanic Technology*, 25, pp. 368–384, <https://doi.org/10.1175/2007JTECHA982.1>.
- Wang, X. *et al.*, 2010: New techniques for the detection and adjustment of shifts in daily precipitation data series. *Journal of Applied Meteorology and Climatology*, 49, pp. 2416–2436, <https://doi.org/10.1175/2010JAMC2376.1>.
- Wang, X.L. et Y. Feng, 2013: *RHtestsV4 User Manual* (RHtestsV4 Manuel de l'utilisateur). Climate Data and Analysis Section – Environment and Climate Change Canada. Publié en ligne (en anglais) en juillet 2013, <https://github.com/ECCC-CDAS>.
- Wang, X.L., Y. Feng et L.A. Vincent, 2013: Observed changes in one-in-20 year extremes of Canadian surface air temperatures. *Atmosphere-Ocean*, 52, pp. 222–231, <https://doi.org/10.1080/07055900.2013.818526>.
- Wang, X.L. *et al.*, 2017: Adjusted daily rainfall and snowfall data for Canada. *Atmosphere-Ocean*, 55:3, pp. 155–168, <https://doi.org/10.1080/07055900.2017.1342163>.
- Wen, Q. H., X.L. Wang et A. Wong, 2011: A hybrid-domain approach to modeling climate data time series. *Journal of Geophysical Research – Atmospheres*, 116, D18112, <https://doi.org/10.1029/2011JD015850>.
- Willett, K.M. *et al.*, 2014: A framework for benchmarking of homogenization algorithm performance on the global scale. *Geoscientific Instrumentation, Methods and Data Systems*, 3, pp. 187–200, <https://gi.copernicus.org/articles/3/>.
- Winkler, P., 2009: Revision and necessary correction of the long-term temperature series of Hohenpeissenberg, 1781–2006. *Theoretical and Applied Climatology*, 98, 259–268, <https://doi.org/10.1007/s00704-009-0108-y>.
- Xu, W., Q. Li, P. Jones, X. L. Wang, B. Trewin, S. Yang, C. Zhu, G. Ren, P. Zhai, J. Wang, L. Vincent, A. Dai, Y. Gao, Y. Ding, 2017: A new integrated and homogenized global monthly land surface air temperature dataset for the period since 1900. *Climate Dynamics*, 50, 2513–2536. <https://doi.org/10.1007/s00382-017-3755-1>.
- Xu W. *et al.*, 2013: Homogenization of Chinese daily surface air temperatures and analysis of trends in the extreme temperature indices. *Journal of Geophysical Research – Atmospheres*, 118(17):9708–9720, <https://doi.org/10.1002/jgrd.50791>.
- Yang, S., X.L. Wang et M. Wild, 2018: Homogenization and trend analysis of the 1958-2016 in-situ surface solar radiation records in China. *Journal of Climate*, 31, 4529–4541, <https://doi.org/10.1175/JCLI-D-17-0891.1>.
- Yosef, Y., E. Aguilar and P. Alpert, 2018. Detecting and adjusting artificial biases of long-term temperature records in Israel. *International Journal of Climatology*, 38(8), 3273–3289, <https://doi.org/10.1002/joc.5500>.
- Zhang, L. *et al.*, 2014: Effect of data homogenization on estimate of temperature trend: a case of Huairou station in Beijing Municipality. *Theoretical and Applied Climatology*, 115, pp. 365–373, <https://doi.org/10.1007/s00704-013-0894-0>.

## AUTRES LECTURES

- Alexandersson, H. et A. Moberg, 1997: Homogenization of Swedish temperature data. Part I: Homogeneity test for linear trends. *International Journal of Climatology*, 17, pp. 25–34, [https://doi.org/10.1002/\(SICI\)1097-0088\(199701\)17:1<25::AID-JOC103>3.0.CO;2-J](https://doi.org/10.1002/(SICI)1097-0088(199701)17:1<25::AID-JOC103>3.0.CO;2-J).
- Auchmann, R. et S. Brönnimann, 2012: A physics-based correction model for homogenizing sub-daily temperature series. *Journal of Geophysical Research – Atmospheres*, 117: D17119, <https://doi.org/10.1029/2012JD018067>.
- Brohan, P. *et al.*, 2006: Uncertainty estimates in regional and global observed temperature changes: A new dataset from 1850. *Journal of Geophysical Research – Atmospheres*, 111, D12106, <https://doi.org/10.1029/2005JD006548>.
- Domonkos, P., 2011: Adapted Caussinus-Mestre algorithm for networks of temperature series (ACMANT). *International Journal of Geosciences*, 2, pp. 293–309. <https://doi.org/10.4236/ijg.2011.23032>.
- Hawkins, D.M., 1972: On the choice of segments in piecewise approximation. *Journal of the Institute of Mathematics and its Applications*, 9, pp. 250–256.
- Killick, R., 2016: *Benchmarking the Performance of Homogenisation Algorithms on Daily Temperature Data*. Thèse de doctorat, Département de mathématiques, Université d'Exeter, Royaume-Uni. <http://hdl.handle.net/10871/23095>.



- Maraun, D. *et al.*, 2010: Precipitation downscaling under climate change: Recent developments to bridge the gap between dynamical models and the end user. *Reviews of Geophysics*, 48, RG3003, <https://doi.org/10.1029/2009RG000314>.
- Maraun, D., 2013: Bias correction, quantile mapping, and downscaling: Revisiting the inflation issue. *Journal of Climate* 26:6, pp. 2137–2143, <https://doi.org/10.1175/JCLI-D-12-00821.1>.
- Morozova, A.L. et M.A. Valente, 2012: Homogenization of Portuguese long-term temperature data series: Lisbon, Coimbra and Porto. *Earth System Science Data*, 4, 187–213, <https://doi.org/10.5194/essd-4-187-2012>.
- Organisation météorologique mondiale (OMM), 1986: *Guidelines on the quality control of surface climatological data* (WMO/TD-No. 111, WCP-85). Genève.
- , 1993: *Guide du système mondial de traitement des données* (OMM-N° 305). Genève.
- , 2003: *Guidelines on climate metadata and homogenization* (WMO/TD-No. 1186, WCDMP No. 53). Genève.
- , 2004: *Guidelines on climate data rescue* (WMO/TD-No. 1210, WCDMP-No. 55). Genève.
- , 2016: *Directives sur les bonnes pratiques en matière de sauvetage des données climatologiques* (OMM-N° 1182). Genève.
- Rienzner, M. et C. Gandolfi, 2013: A procedure for the detection of undocumented multiple abrupt changes in the mean value of daily temperature time series of a regional network. *International Journal of Climatology*, 33, pp. 1107–1120, <https://doi.org/10.1002/joc.3496>.
- Rustemeier, E. *et al.*, 2017: HOMPR Europe – A gridded precipitation dataset from European homogenized time series. In *9th Seminar for Homogenization and Quality Control in Climatological Databases and 4th Conference on Spatial Interpolation Techniques in Climatology and Meteorology*, Budapest, Hungary, 3–7 April 2017 (dir. publ.: T. Szentimrey, L. Hoffmann et M. Lakatos). Budapest, Service météorologique hongrois (OMSZ).
- Von Storch, H., 1999: On the use of “inflation” in statistical downscaling. *Journal of Climate*, 12, 3505–3506, [https://doi.org/10.1175/1520-0442\(1999\)012<3505:OTUOII>2.0.CO;2](https://doi.org/10.1175/1520-0442(1999)012<3505:OTUOII>2.0.CO;2).
- Yosef, Y., I. Osetinsky-Tzidaki et A. Furshpan, 2015: Homogenization of monthly temperature series in Israel – An integrated approach for optimal break-points detection. *Eighth Seminar for Homogenization and Quality Control in Climatological Databases and Third Conference on Spatial Interpolation Techniques in Climatology and Meteorology*, Budapest, Hungary, 12–16 May 2014 (WCDMP-No. 84). Genève, OMM.
-

Pour de plus amples informations, veuillez vous adresser à:

**Organisation météorologique mondiale**

7 bis, avenue de la Paix – Case postale 2300 – CH 1211 Genève 2 – Suisse

**Bureau de la communication stratégique**

Tél.: +41 (0) 22 730 87 40/83 14 – Fax: +41 (0) 22 730 80 27

Courriel: [communications@wmo.int](mailto:communications@wmo.int)

[public.wmo.int](http://public.wmo.int)