

# RHtestsV4

## Manuel de l'utilisateur

Xiaolan L. Wang et Yang Feng

Division de la recherche climatique  
Direction des sciences et technologies atmosphériques  
Direction générale des sciences et de la technologie  
Environnement Canada  
Toronto, Ontario, Canada

Publié en ligne (en anglais) à  
<http://etccdi.pacificclimate.org/software.shtml>  
20 Juillet 2013

## **Table des matières**

0. Guide des citations

1. Introduction

2. Format des données d'entrée pour RHtestsV4

3. Comment utiliser les fonctions de RHtestsV4

    3.1 Démarrage rapide avec l'interface graphique (GUI)

    3.2 Cas « avec série de référence »

    3.3 Cas « sans série de référence »

Références

Remerciements

## Guide des citations

Les utilisateurs de ce logiciel doivent citer dans leurs publications ce manuel de l'utilisateur:

Wang, X. L. and Y. Feng, published online July 2013: *RHtestsV4 User Manual*. Climate Research Division, Atmospheric Science and Technology Directorate, Science and Technology Branch, Environment Canada. 28 pp. [Disponible en ligne à <http://etccdi.pacificclimate.org/software.shtml>],

ainsi que les publications suivantes pour les méthodes mises en œuvre dans ce forfait:

- Veuillez citer ces deux articles si vous faites les tests d'homogénéité avec la série de référence :  
Wang, X. L., 2008: Accounting for autocorrelation in detecting mean-shifts in climate data series using the penalized maximal  $t$  or  $F$  test. *J. Appl. Meteor. Climatol.*, 47, 2423-2444.  
Wang, X. L., Q. H. Wen, and Y. Wu, 2007: Penalized maximal  $t$  test for detecting undocumented mean change in climate data series. *J. Appl. Meteor. Climatol.*, 46 (No. 6), 916-931. DOI:10.1175/JAM2504.1
- Veuillez citer ces deux articles si vous faites les tests d'homogénéité sans la série de référence :  
Wang, X. L., 2008a: Accounting for autocorrelation in detecting mean-shifts in climate data series using the penalized maximal  $t$  or  $F$  test. *J. Appl. Meteor. Climatol.*, 47, 2423-2444.  
Wang, X. L., 2008b: Penalized maximal  $F$ -test for detecting undocumented mean-shifts without trend-change. *J. Atmos. Oceanic Tech.*, 25 (No. 3), 368-384. DOI:10.1175/2007/JTECHA982.1.
- Veuillez citer ces deux articles si vous utilisez les données QM ajustées dans la sortie des données homogénéisées (c'est-à-dire que vous utilisez l'algorithme de réglage Quantile-Matching):  
Wang, X. L., H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.* 49 (No. 12), 2416-2436. DOI: 10.1175/2010JAMC2376.1  
Vincent, L. A., X. L. Wang, E. J. Milewska, H. Wan, Y. Feng, and V. Swail, 2012: A Second Generation of Homogenized Canadian Monthly Surface Air Temperature for Climate Trend Analysis, *JGR-Atmospheres*, 117, D18110, doi:10.1029/2012JD017859.

### 1. Introduction

Le progiciel RHtestsV4 est le progiciel RHtestsV3 qui inclut l'ajout de l'ajustement QM qui est estimé à partir d'une station de référence (dans les cas où la détection des points de changement est faite avec une station de référence; Vincent et al. 2012). Le progiciel RHtestsV3 est une version prolongée du progiciel RHtestV2. Celle-ci inclut : (1) l'ajout de l'ajustement QM (Quantile-Match) décrit dans Wang et al. 2010 en plus de l'ajustement-moyen fourni dans les versions précédentes ; (2) choix d'utiliser tout le segment ou une partie du segment avant et après le saut pour estimer les ajustements QM; (3) choix du segment pour lequel la série de base doit être ajustée (désigné sous le nom

du segment de base); (4) choix du niveau nominal de confiance utilisé pour effectuer le test; et (5) toutes les fonctions sont maintenant disponibles dans le mode de l'interface graphique GUI.

L'objectif de l'ajustement QM est d'ajuster la série de sorte que la distribution empirique de tous les segments de la série de base tendance-enlevée s'assortissent ; la valeur de l'ajustement dépend alors de la fréquence empirique des valeurs à ajuster (c.-à-d. elle varie d'une donnée à l'autre dans le même segment, selon sa fréquence empirique correspondante). En conséquence, la forme de la distribution est souvent ajustée (comprenant, mais non limité à, l'ajustement de la moyenne), bien que les tests soient censés de détecter les sauts dans les moyennes (ainsi, un changement de la distribution *sans variation dans la moyenne* pourrait être non détecté); et l'ajustement QM tient compte de la saisonnalité de la changement (p. ex., c'est possible que les températures de l'hiver et de l'été sont ajustées différemment à cause qu'ils appartiennent à des quantiles différents de la distribution). Aussi, le cycle annuel, l'autocorrélation de retard de 1, et la tendance linéaire de la série de base ont été estimés tout en expliquant tous les sauts identifiés (Wang 2008a) ; et la tendance prévue pour la série de base est préservée dans les ajustements QM lorsqu'ils sont estimés à partir d'une série de référence. Lorsque c'est possible, les ajustements QM devraient être estimés utilisant une série de référence qui est homogène sur une période comprenant le saut dans la série de base qui doit être ajustée.

Ce progiciel peut être utilisé pour détecter et ajuster des points de changement multiples (des sauts) qui peuvent exister dans une série de données pouvant comporter des erreurs autorégressives du premier ordre [mais les séries de précipitation doivent être traitées avec le logiciel RHtests\_dlyPrpc (Wang et al. 2010)]. Ce logiciel est basé sur le test  $t$  maxima avec pénalité (Wang et al. 2007) et le test  $F$  maxima avec pénalité (Wang 2008b), imbriqués dans un algorithme de test récursif (Wang 2008a), avec un facteur d'autocorrélation avec retard de 1 (si significatif) appliqué à la série temporelle étant prise en compte. La série temporelle testée peut présenter une tendance nulle ou linéaire pendant toute la durée enregistrée. Le problème de la répartition non uniforme des taux de fausse alarme et de la puissance de détection est aussi grandement réduit au moyen de fonctions de pénalité empiriques. Par conséquent, RHtestsV4 (et RHtestsV3 et RHtestV2) présentent des améliorations importantes par rapport à la version RHtestV0.9, qui ne tenait compte d'aucune autocorrélation et qui ne résolvait pas le problème de la distribution inégale du taux de fausse alarme et de la puissance de détection (Wang et al. 2007, Wang 2008b). Une série temporelle homogène bien corrélée avec la série de base peut être utilisée comme série de référence. Aussi, RHtestsV4 permet la détection de points de changement lorsqu'on ne dispose pas de série de référence homogène. Mais les résultats sont moins sûrs et ont besoin plus d'analyse. **On ne devrait pas faire l'homogénéisation de données d'une façon automatique sans une série de référence (c'est-à-dire sans plus d'analyse manuelle des résultats statistiques).**

Ce simple manuel constitue un aide-mémoire pour l'utilisation des fonctions du progiciel RHtestsV4 (et aussi pour l'utilisation des fonctions équivalentes en FORTRAN, que l'on pourra obtenir en faisant une demande en anglais auprès de [Xiaolan.Wang@ec.gc.ca](mailto:Xiaolan.Wang@ec.gc.ca)). Il

est supposé que les utilisateurs sont familiers avec l'utilisation de R (débuter une session R et y mettre fin, et appeler une fonction R).

## 2. Format des données d'entrée pour RHtestsV4

- Les fonctions de RHtestsV4 peuvent traiter des séries annuelles, mensuelles ou quotidiennes d'erreurs gaussiennes (veuillez noter que la procédure RHtests\_dlyPrpc doit être utilisée pour l'homogénéisation des séries de précipitation quotidienne qui sont typiquement non-gaussiennes); **il est cependant correct d'utiliser les fonctions RHtestsV4 sur une transformée-log des précipitations totales mensuelles et annuelles.** Chaque série de données d'entrée doit être stockée dans un fichier distinct (p. ex., un fichier nommé Example.dat), dans lequel les trois premières colonnes représentent les dates d'observation (AAAA pour l'année du calendrier, MM pour le mois et JJ pour le jour) et la quatrième colonne, les valeurs des données d'observation (ou code de valeur absente). Il est à noter que pour les séries mensuelles, JJ=00, et pour les séries annuelles, MM=00 et JJ=00. Par exemple :

(Série quotidienne)				OU	(Série mensuelle)			
...					...			
1994	1	27	8.1		1967	7	00	1015.70
1994	1	28	5.3		1967	8	00	1015.95
1994	1	29	4.9		1967	9	00	1016.10
1994	1	30	4.9		1967	10	00	-999.99
1994	1	31	4.0		1967	11	00	1010.71
1994	2	1	3.9		1967	12	00	1011.58
1994	2	2	7.2		1968	1	00	1009.37
1994	2	3	8.7		1968	2	00	1003.07
1994	2	4	6.3		1968	3	00	1011.94
1994	2	5	-999.		1968	4	00	1014.74
1994	2	6	-999.		1968	5	00	1009.59
1994	2	7	-999.		1968	6	00	1011.77
1994	2	8	-999.		1968	7	00	1014.35
1994	2	9	9.0		1968	8	00	1010.87
1994	2	10	6.0		1968	9	00	1016.45
...					...			

Les dates des données d'entrée **doivent être consécutives** et dans l'ordre du calendrier. Autrement, le programme se fermera en produisant un message d'erreur contenant la première date où l'erreur de données s'est produite. Par exemple : les quatre lignes du 5 au 8 février 1994 de l'exemple de série quotidienne ci-dessus doivent être incluses dans le fichier d'entrée de données. Elles ne doivent pas être supprimées en raison de leurs données manquantes. La ligne du 10 février 1994 ne doit pas être avant le 9 février 1994, etc.

- L'exigence ci-dessus s'applique à la série de base et à la série de référence, si elles sont utilisées. Les séries de base et de référence peuvent comporter des dates pour diverses périodes (il n'est pas nécessaire qu'elles soient de même longueur), mais dans ce cas, seules les périodes communes à la série de base et à celle de référence sont testées/analysées. Il est à noter qu'il peut manquer des valeurs à certaines dates/heures dans les séries de base et de référence, mais alors il **faut** que le même code soit utilisé pour indiquer les valeurs absentes. Toutes les dates

- et heures correspondant aux valeurs manquantes, **que ce soit** dans la série de base, **ou** dans la série de référence, ou dans les deux séries, sont alors exclues de l'analyse.
- Il existe diverses manières pour construire une série de référence (p. ex., on pourra appliquer divers poids à diverses stations pour composer une série de référence, ou on pourra utiliser une série provenant d'une seule station comme série de référence). Dans ce cas, pour chaque série de base, il est supposé qu'une seule série de référence a été construite par vous (l'utilisateur), selon votre choix. **L'homogénéité des données de la série de référence peut être vérifiée en appelant la fonction *FindU*** (qui n'a pas besoin de série de référence). Les sauts significatifs de la série de référence doivent être ajustés avant qu'elle ne soit utilisée pour vérifier l'homogénéité de la série de base.
  - Il est recommandé de tester les séries mensuelles avant de tester les séries journalières car les séries journalières ont une plus grande variabilité et il est ainsi plus difficile de retrouver les points de changement dans ces séries. Les résultats obtenus des séries mensuelles peuvent être utilisés subjectivement pour l'analyse des séries journalières. Il est aussi recommandé d'effectuer la normalisation logarithmique d'une série de précipitations totales mensuelles avant d'utiliser cette dernière comme série de données d'entrée, car les tests effectués par ce logiciel supposent que la série de données comportera des erreurs gaussiennes (ce qui est fait dans `RHtests_dlyPrcp`).

### 3. Comment utiliser les fonctions de `RHtestsV4`

Le progiciel `RHtestsV4` comporte six fonctions permettant de détecter dans une série de données des sauts artificiels et d'effectuer les ajustements requis, avec ou sans l'utilisation d'une série de référence.

Commencez par entrer `source("RHtestsV4.r")` à l'invite R ("`>`") pour charger les fonctions de `RHtestsV4` dans R.

**Voici en bref les étapes à suivre :** (voir les sections 3.2 et 3.3 pour plus de détails)

- 1) Appelez la fonction `FindU` (ou `FindU.wRef`) avec la liste appropriée des paramètres (voir section 3.2 ou 3.3).
- 2) Allez à l'étape 5 si vous n'avez pas de métadonnées. Sinon, appelez `FindUD` (ou `FindUD.wRef`) avec la liste appropriée des paramètres (voir section 3.3).
- 3) Modifiez le fichier résultant `*_mCs.txt` file (liste de points de changement identifiés jusqu'à présent qui sont dans le répertoire où les données sont testées), au besoin, pour intégrer des informations de métadonnées aux résultats. (Dans ce cas, l'astérisque (\*) représente un préfixe précisé par l'utilisateur comme nom des fichiers de sortie.)
- 4) Appelez la fonction `StepSize` (ou `StepSize.wRef`) avec la liste appropriée des paramètres (voir section 3.2 ou 3.3) pour évaluer l'importance et l'ordre de grandeur des points de changement retenus.

- 5) Analysez la plus récente version du fichier \*\_mCs.txt et supprimez le plus petit saut si vous déterminez statistiquement ou subjectivement qu'il n'est pas significatif. Appelez ensuite la fonction *StepSize* (ou *StepSize.wRef*) une autre fois pour réévaluer l'importance des autres points de changement. Recommencez procédure (5) jusqu'à ce que tous les points de changement de la liste soient estimés être significatifs.

**Spécifiquement, la procédure générale est la suivante :**

- (1) Appelez la fonction *FindU* ou *FindU.wRef* pour détecter tous les points de changement qui pourraient être significatifs au niveau nominal même sans soutien de métadonnées (ils sont nommés points de changement de type 1). Si aucun point de changement significatif n'a été identifié à cette étape, les séries temporelles testées peuvent être déclarées homogènes et il n'est pas nécessaire de tester cette série plus avant.
- (2) **Allez au point (5) si aucune métadonnée n'est disponible ou si vous voulez détecter seulement les points de changement significatifs sans soutien de métadonnées, c.-à-d. les points de changement de type 1.**  
Sinon, appelez la fonction *FindUD* ou *FindUD.wRef*. Les points de changement additionnels se nomment points de changement de type 0 (ces points de changement peuvent être significatifs **seulement si** ils sont soutenus par des métadonnées fiables). Cette étape peut-être utilisée pour aider l'investigation des métadonnées qui comprend aussi les points de changement de type 0). Cependant, il est suffisant d'appliquer ces fonctions sur des séries mensuelles ou annuelles, et il n'est pas nécessaire, et cela prendra trop de temps, de les appliquer sur les données quotidiennes. Lorsqu'on analyse les séries des données quotidiennes, on devrait appliquer ces fonctions sur les séries mensuelles correspondantes.
- (3) Examinez les métadonnées pour voir s'il y a eu des événements **précis** aux dates des points de changement, ou **à leur proximité**, qui auraient pu provoquer des sauts. Conservez seulement les points de changement de type 0 qui sont soutenus par des métadonnées, ainsi que tous les points de changement de type 1 identifiés au point (1). La date d'un point de changement peut être changée à la date de changement documentée obtenue des métadonnées si on est certain de la cause du changement; on peut aussi changer le type de point de changement de 1 à 0 si on constate qu'un point de changement de type 1 dispose d'un soutien fiable de métadonnées.
- (4) Appelez la fonction *StepSize* (ou *StepSize.wRef*) pour évaluer le degré de signification et l'ordre de grandeur des points de changement restants (énumérés dans la dernière version du fichier \*\_mCs.txt).
- (5) Analysez la plus récente version du fichier \*\_mCs.txt et supprimez le point de changement le moins significatif s'il est déterminé qu'il n'est pas significatif au niveau nominal. Appelez ensuite la fonction *StepSize* (ou *StepSize.wRef*) pour réévaluer le degré de signification des points de changement restants. Recommencez cette procédure (5) jusqu'à ce que tous les points de changement retenus soient significatifs.

Dans le mode de l'interface graphique (GUI – section 3.1), le fichier final est dans le sous répertoire OUTPUT du répertoire des données (c.-à-d. où se trouvent les données à tester); les données et les répertoires sont indiqués dans la fenêtre GUI. Dans la ligne d'instructions (section 3.2 ou 3.3), l'utilisateur peut spécifier le répertoire pour emmagasiner les fichiers en l'indiquant dans la chaîne des paramètres de sortie (voir section 3.2 ou 3.3)

### 3.1 Démarrage rapide avec l'interface graphique (GUI)

Les fonctions *FindU*, *FindUD*, et *StepSize* sont basé sur le test de  $F$  maxima (PMF) avec pénalité (Wang 2008a et 2008b), qui permet à une série temporelle testée d'avoir une tendance nulle ou linéaire pendant toute la durée enregistrée (c.-à-d. qu'il n'y a pas de changement dans la tendance; voir Wang 2003), avec le cycle annuel, la tendance linéaire, et l'autocorrélation de retard de 1 de la série de base estimés en utilisant des procédures itératives tout en expliquant tous les sauts identifiés (Wang 2008a). Aucune série de référence n'est utilisée ici dans ces fonctions; c'est la série de base qui est testée dans ce cas-ci. Veuillez vous référer à la section 3.3 pour plus de détails au sujet de ces fonctions.

Les fonctions *FindU.wref*, *FindUD.wref*, et *StepSize.wref* sont basé sur le test de  $t$  maxima (PMT) avec pénalité (Wang 2008a, Wang et al. 2007), qui assume une série temporelle de base ayant une tendance nulle et des erreurs gaussiennes. Une série de référence doit être utilisée pour exécuter les fonctions *\*.wref*. La différence entre les séries de base et référence est testée pour identifier les positions et la signifiante des points de changement, mais un modèle de régression multi-phase (MPR) avec tendance est aussi appliqué aux anomalies (relatives au cycle annuel) de la série de base pour obtenir les estimés finaux de la grandeur des changements (voir l'appendice dans Wang 2008a pour plus de détails). Durant l'application de MPR, le cycle annuel, la tendance linéaire et l'autocorrélation de retard de 1 sont estimés en utilisant des procédures itératives tout en expliquant tous les sauts identifiés (Wang 2008a). Veuillez vous référer à la section 3.3 pour plus de détails au sujet de ces fonctions.

La fonction *QMadj\_GaussDLY* est utilisée pour appliquer l'ajustement QM (Wang et al. 2010) sur les températures quotidiennes (ou sur toutes données gaussiennes en général) pour être ajustées pour une liste de points de changement significatifs identifiés (p. ex. après l'application d'une ou de plusieurs fonctions décrites ci-dessus). **Cette fonction n'est pas appropriée pour ajuster les précipitations quotidiennes.** Veuillez appeler la fonction *StepSize\_dlyPrcp* dans le logiciel RHtests\_dlyPrcp (aussi disponible sur le site web) pour ajuster les précipitations quotidiennes pour les points de changement identifiés.

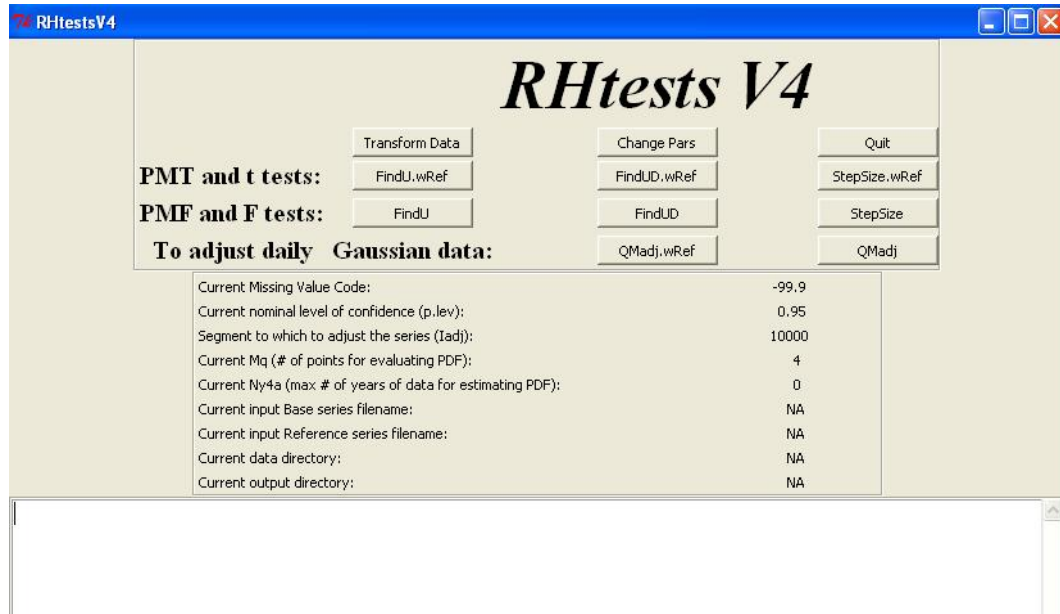
Dans ce mode à interface graphique (GUI) simple, le préfixe du nom du fichier d'entrée de données est utilisé comme préfixe du nom des fichiers de sortie. Par exemple, dans le cas où une série de référence n'est pas utilisée, si *Example.dat* est le nom du fichier d'entrée de données, le fichier de sortie se nommera *Example\_\*.\**; lorsqu'une série de



référence est utilisée, par exemple, Base.dat et Ref.dat, données de base et de référence, les fichiers de sortie seront nommés Base\_Ref\_\*. \*.

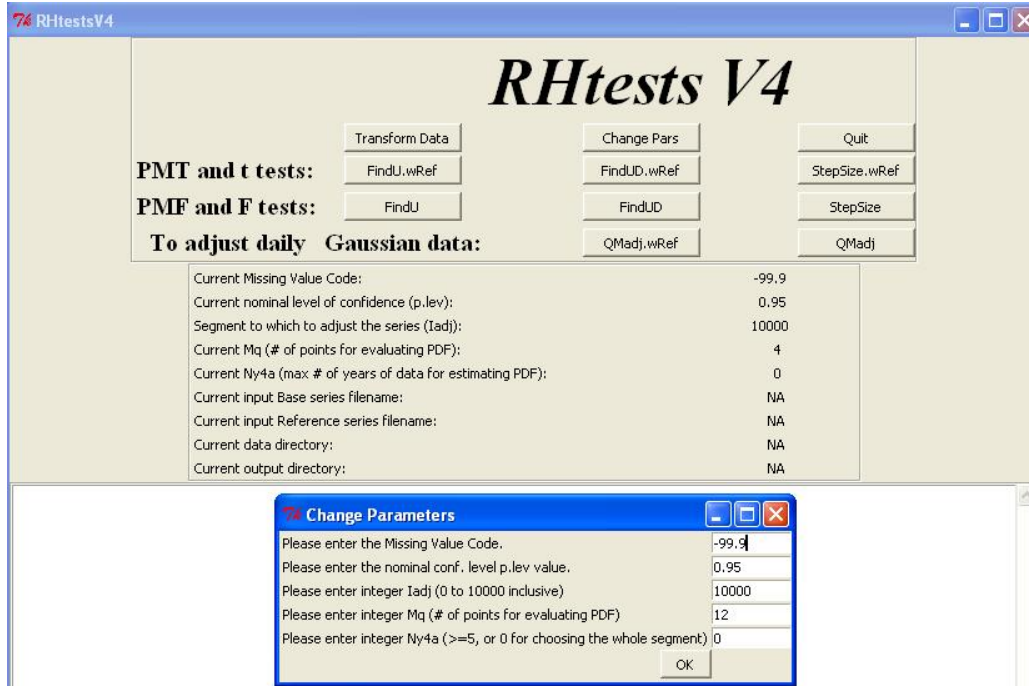
Plus spécifiquement, la procédure est la suivante :

- (1) Pour lancer une session de l'interface graphique, entrez **StartGUI()** après avoir entré *source("RHtestsV4.r")* à l'invite R. La fenêtre suivante s'ouvre.

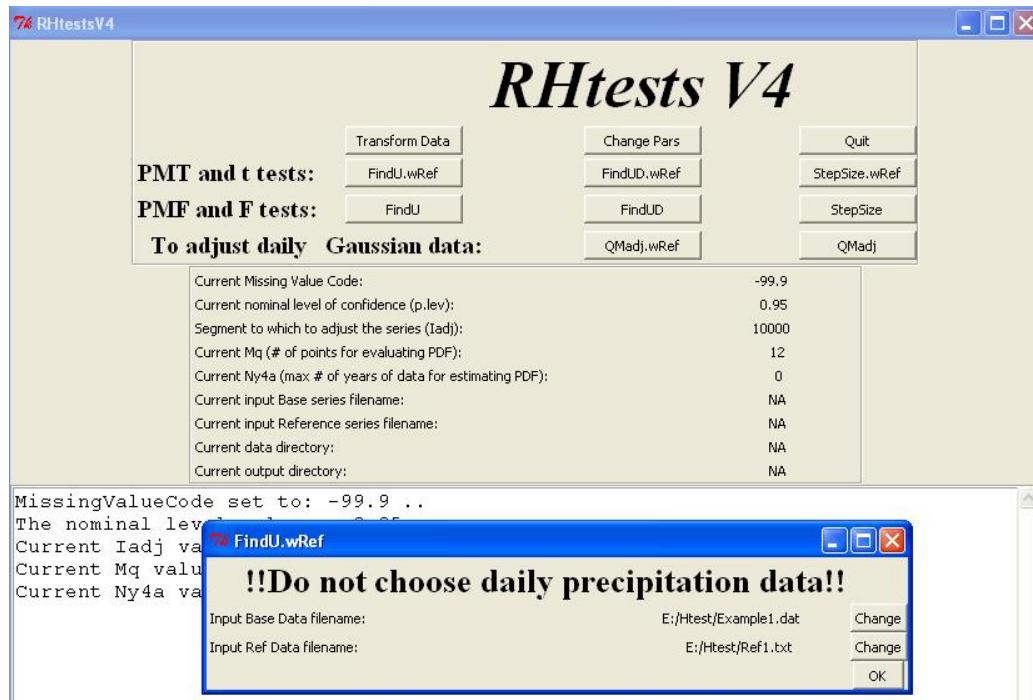


- (2) **Les usagers n'utilisant pas RClimDex doivent sauter cette procédure (2).** Pour convertir les séries de données quotidiennes dans le format standard de RClimDex en une série de moyenne mensuelle en format standard RHtestsV4, cliquez sur le bouton **Transform Data**, sélectionnez le fichier de série de données quotidiennes à convertir et cliquez sur le bouton **Open**. Cela produit neuf fichiers : \*\_tmaxMLY.txt, \*\_tminMLY.txt, \*\_prcpMLY.txt, \*\_tminDLY.txt, \*\_tmaxDLY.txt, \*\_prcpDLY.txt, \*\_prcpMLY1mm.txt, \*\_LogprcpMLY.txt, and \*\_LogprcpMLY1mm.txt (l'astérisque (\*) représente le préfixe du nom du fichier d'entrée, DLY et MLY pour les données journalières et mensuelles respectivement, et MLY1mm pour les totaux mensuels des précipitations journalières  $\geq 1$ mm).
- (3) Cette étape n'est pas requise si les données sont converties au moyen de la fonction **Transform Data**. Cliquez sur le bouton **ChangePars** pour choisir les paramètres suivants : (a) le code de valeur absente utilisé dans la série de données à tester, p. ex., « -999.0 » dans la fenêtre ci-dessous (il est à noter que le code entré ici **doit être exactement le même** que celui qui est utilisé dans les données; p. ex., « -999. » et « -999.0 » sont différents; on ne peut pas entrer « -999. » au lieu de « -999.0 » lorsque « -999.0 » est utilisé dans la série de données d'entrée; cela produirait des données erronées); (b) le niveau de confiance nominal pour lequel on conduit le test; (c) les segment de base (pour lequel on ajuste la série); (d) le nombre de points (Mq) pour lesquels la distribution de probabilité empirique (PDF) est estimée en utilisant les ajustements QM; et (e) le nombre maximal

d'années de données immédiatement avant ou après le point de rupture qui est utilisé pour estimer le PDF ( $Ny4a = 0$  pour choisir le segment au complet). Ensuite veuillez cliquer sur le bouton OK pour accepter les valeurs des paramètres affichés dans la fenêtre.



- (4) Si vous voulez utiliser une série de référence, veuillez cliquer sur le bouton **FindU.wref** qui fera afficher une fenêtre, puis cliquer sur le bouton **Change** pour choisir, un par un, le fichier de base et de référence (p. ex. Example1.dat et Ref1.txt ci-dessous; NA signifie que le fichier n'a pas été choisi), et cliquer sur **OK** pour exécuter le test PMT. Cela produit les fichiers suivants dans le répertoire de sortie : Example\_Ref1\_1Cs.txt, Example\_Ref1\_Ustat.txt, Example\_Ref1\_U.dat et Example\_Ref1\_U.pdf (voir section 3.2 pour avoir une description du contenu de ces fichiers). Une copie du premier fichier est aussi stockée dans le fichier Example\_Ref1\_mCs.txt du répertoire de sortie, qui énumère tous les points de changement qui peuvent être significatifs au niveau nominal, même sans soutien de métadonnées (c.-à-d. les points de changement de type 1).



Si vous n'utilisez pas de station de référence, veuillez cliquer sur le bouton **FindU**, qui fera afficher une fenêtre (non montrée ici); sélectionnez la série de données (par exemple, Example1.dat) à tester et cliquez sur le bouton **Open**. Cela produit les fichiers suivants dans le répertoire de sortie : Example1\_1Cs.txt, Example1\_Ustat.txt, Example1\_U.dat et Example1\_U.pdf (voir section 3.3 pour avoir une description du contenu de ces fichiers). Une copie du premier fichier est aussi stockée dans le fichier Example\_mCs.txt du répertoire de sortie, qui énumère tous les points de changement qui peuvent être significatifs au niveau nominal, même sans soutien de métadonnées (c.-à-d. les points de changement de type 1).

Un exemple du fichier \*1Cs.txt ressemble à ceci:

```

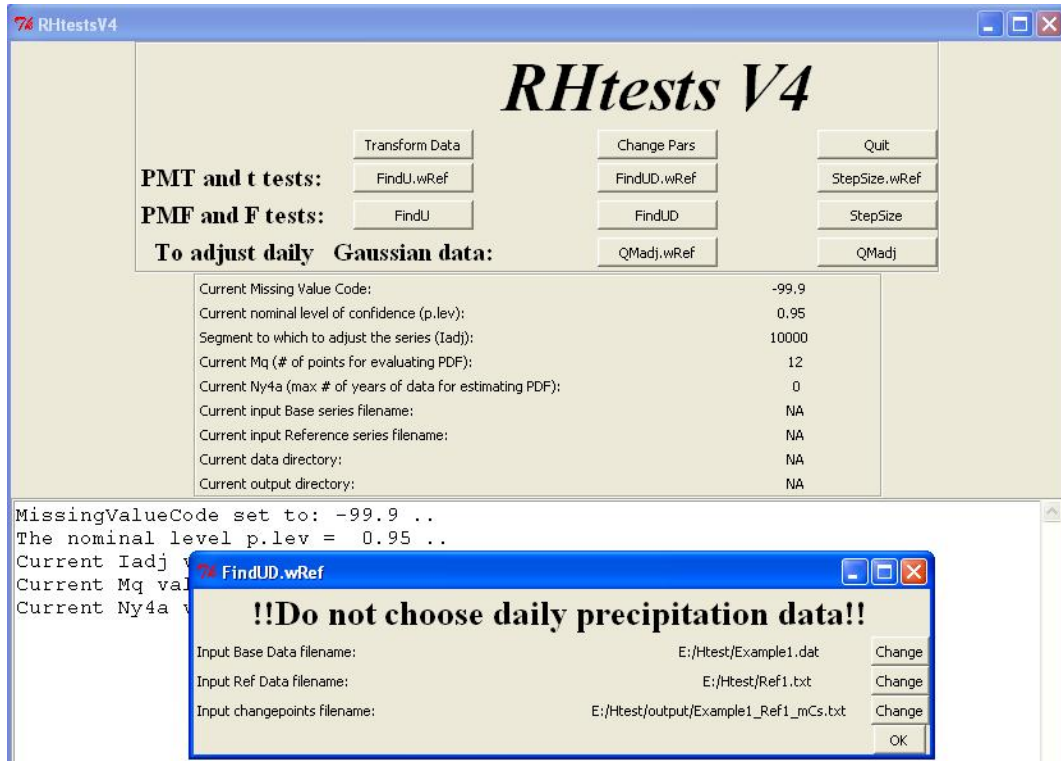
10 changepoints in Series ...Example2.dat ...
1 Yes 19600200 (1.0000-1.0000) 0.950 5.3553 ( 3.0635- 3.5582)
1 Yes 19650700 (0.9999-0.9999) 0.950 4.1634 ( 3.0977- 3.6009)
1 Yes 19800300 (1.0000-1.0000) 0.950 5.2305 ( 3.0905- 3.5927)
1 Yes 19820800 (1.0000-1.0000) 0.950 4.6415 ( 3.0237- 3.5000)
1 Yes 19850200 (1.0000-1.0000) 0.950 4.7195 ( 3.0005- 3.4712)
1 Yes 19851100 (0.9999-0.9999) 0.950 3.8607 ( 3.0457- 3.5283)
1 Yes 19930200 (1.0000-1.0000) 0.950 9.1678 ( 3.0469- 3.5342)
1 Yes 19940600 (1.0000-1.0000) 0.950 5.9064 ( 2.9845- 3.4552)
1 ? 19950900 (0.9965-0.9971) 0.950 3.0340 ( 2.9929- 3.4612)
1 Yes 19970400 (1.0000-1.0000) 0.950 8.2999 ( 3.0497- 3.5398)

```

Ici, la première colonne (les 1) représente les points de changement de type 1. La deuxième colonne indique si le point de changement est significatif ou non pour le type fourni à la première colonne; est-ce que ce point est oui ou peut-être

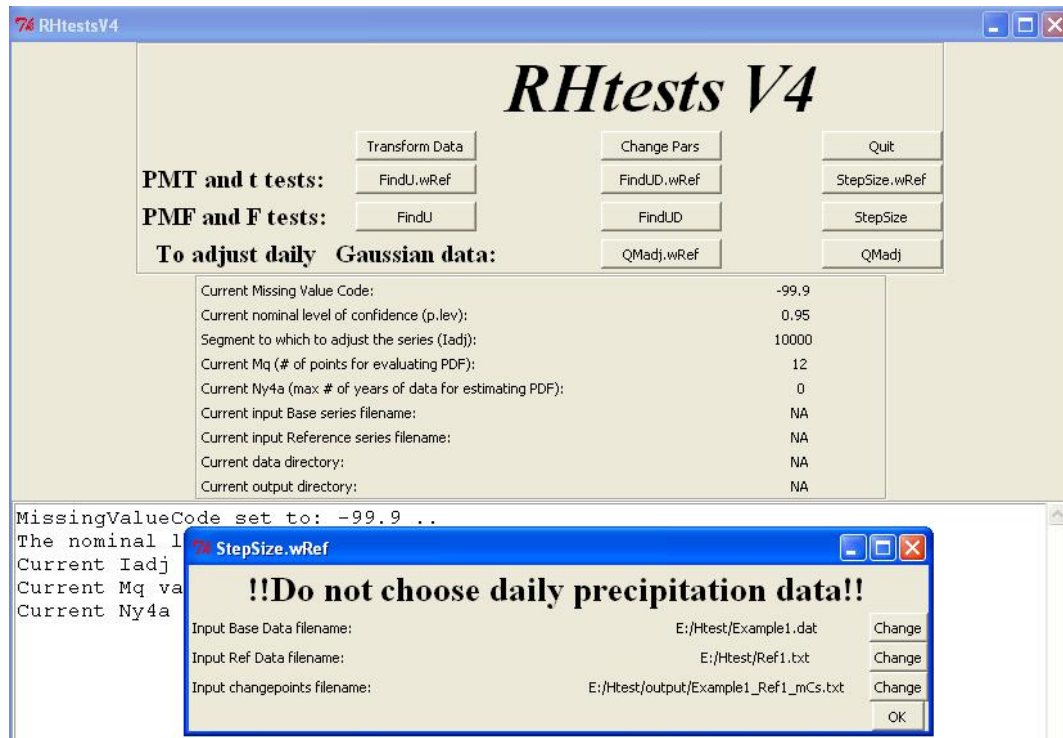
(« Yes » or « ? ») dans le fichier \*\_1Cs.txt, mais dans les autres fichiers \*Cs.txt ils peuvent être : (1) “Yes” (significatif); (2) “No” (pas significatif pour le type fourni dans la première colonne; (3) “?” (peut-être significatif pour le type fourni dans la première colonne, et (4) “YifD” (significatif si documenté, c’est-à-dire supporté par l’historique de la station). La troisième colonne liste les dates des points de changement YYYYMMDD : par exemple 19700100 indique janvier 1970. Les nombres dans la quatrième colonne (entre parenthèses) fournissent l’intervalle de confiance à 95% de la valeur p, qui est estimée assumant que le point de changement est documenté. La valeur p nominale (niveau de confiance) est fournie dans la cinquième colonne. Les trois dernières colonnes fournissent la valeur du test statistique  $PF_{max}$  (ou  $PT_{max}$ ) et l’intervalle de confiance à 95% des centiles de  $PF_{max}$  (or  $PT_{max}$ ) qui correspondent au niveau de confiance. Une copie du fichier OutFile\_1Cs.txt est sauvegardée dans le fichier OutFile\_mC.txt pour des modifications postérieures (pour que l’original reste inchangé).

- (5) **Si vous connaissez des points de changement qui peuvent causer des sauts, ajoutez les dates correspondantes à ces points au fichier Example1\_Ref1\_mC.txt ou Example1\_mC.txt s’ils ne sont pas déjà-là, et allez à la procédure (7) ci-dessous. Si vous n’avez pas de métadonnées, ou si vous voulez seulement détecter les points de changement de type 1, passez aussi à la procédure (7).** Sinon, cliquez sur le bouton FindUD.wref (ou FindUD dans le cas pour lequel il n’y a pas de référence) pour identifier tous les points de changement de type 0 (c.-à-d. les points de changement qui peuvent être significatifs *seulement* s’ils sont soutenus par des métadonnées) pour le fichier d’entrée choisi. La fenêtre ci-dessous sera affichée pour choisir ou confirmer les fichiers d’entrée pour exécuter la fonction FindUD.wref (ou FindUD). Puisque cette étape peut aider à resserrer l’investigation des métadonnées, il est suffisant d’appliquer ces fonctions sur des séries mensuelles ou annuelles, et il n’est pas nécessaire, et cela prendra trop de temps, de les appliquer sur les données quotidiennes. Lorsqu’on analyse les séries des données quotidiennes, on devrait appliquer des fonctions sur les séries mensuelles correspondantes.



Quatre fichiers seront produits dans le répertoire de sortie, p. ex., Example1\_Ref1\_pCs.txt, Example1\_Ref1\_UDstat.txt, Example1\_Ref1\_UD.dat et Example1\_Ref1\_UD.pdf en appelant la fonction **FindUD.wref** avec les fichiers Example1.dat et Ref1.txt comme séries de base et référence (voir section 3.2 pour avoir une description du contenu de ces fichiers); ou Example1\_pCs.txt, Example1\_UDstat.txt, Example1\_UD.dat et Example1\_UD.pdf en appelant la fonction **FindUD** avec le fichier d'entrée Example1.dat (voir section 3.3 pour avoir une description du contenu de ces fichiers). Une copie du fichier Example1\_Ref1\_pCs.txt (ou Example1\_pCs.txt) est aussi stockée dans le fichier Example1\_Ref1\_mCs.txt (ou Example1\_mCs.txt) dans le répertoire de sortie (pour que la version précédente soit mise-à-jour).

- (6) Examinez les métadonnées et supprimez du fichier Example1\_Ref1\_mCs.txt ou Example1\_mCs.txt dans le répertoire de sortie tous les points de changement de type 0 qui ne sont pas soutenus par des métadonnées. Cliquez sur le bouton **StepSize.wref** (ou **StepSize** si la série de référence n'est pas utilisée) pour réévaluer le degré de signification/l'ordre de grandeur des points de changement restants, ce qui produira les fichiers suivants : Example1\_Ref1\_fCs.txt, Example1\_Ref1\_Fstat.txt, Example1\_Ref1\_F.dat et Example1\_Ref1\_F.pdf et le fichier Example1\_Ref1\_mCs.txt (ou Example1\_fCs.txt, Example1\_Fstat.txt, Example1\_F.dat et Example1\_F.pdf et le fichier Example1\_mCs.txt) dans le répertoire de sortie (voir section 3.2 ou 3.3 pour avoir une description du contenu de ces fichiers). Veuillez vérifier les noms des fichiers d'entrée pour vous assurez de ce que vous avez choisi.

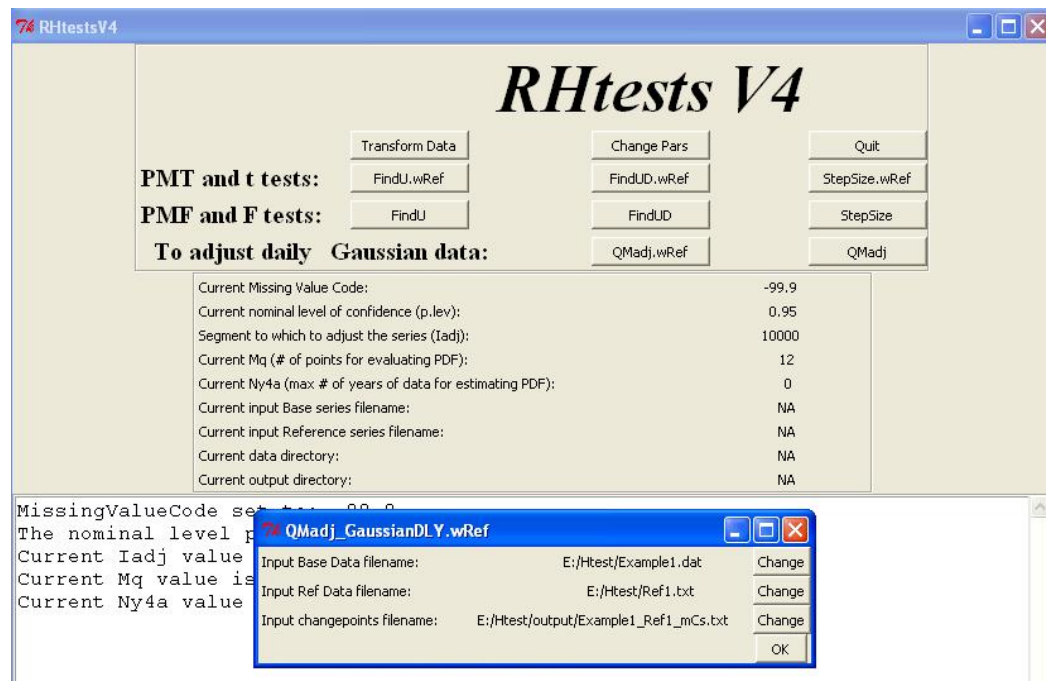


- (7) Analysez les résultats obtenus pour déterminer si le plus petit saut est significatif [voir (T5) dans la section 3.2 ou (F5) dans la section 3.3 pour connaître les détails sur la manière de procéder]. Si vous déterminez qu'il n'est pas significatif, supprimez-le du fichier Example1\_Ref1\_mCs.txt ou Example1\_mCs.txt du répertoire de sortie et cliquez sur le bouton **StepSize.wref** ou **StepSize** pour réévaluer le degré de signification et l'ordre de grandeur des points de changement restants, ce qui mettra à jour (ou produira) les fichiers suivants avec les nouveaux estimés : Example1\_Ref1\_fCs.txt, Example1\_Ref1\_Fstat.txt, Example1\_Ref1\_F.dat et Example1\_Ref1\_F.pdf et Example1\_Ref1\_mCs.txt (ou Example1\_fCs.txt, Example1\_Fstat.txt, Example1\_F.dat et Example1\_F.pdf et Example1\_mCs.txt) dans le répertoire de sortie.
- (8) Recommencez la procédure (7) ci-dessus, jusqu'à ce que tous les points de changement retenus dans le fichier Example1\_Ref1\_mCs.txt (ou Example1\_mCs.txt) soient estimés comme étant significatifs (aucune autre suppression ne sera effectuée). Les quatre fichiers finals de sortie suivants sont alors dans le répertoire de sortie :
- Example\_Ref1\_fCs.txt (Example\_fCs.txt), qui énumère les points de changement identifiés avec leur degré de signification et leurs statistiques;
  - Example\_Ref1\_Fstat.txt (ou Example\_Fstat.txt), qui stocke l'amplitude des sauts estimés et une copie du contenu dans Example\_Ref1\_fCs.txt (ou Example\_fCs.txt);
  - Example\_Ref1\_F.dat (ou Example\_F.dat), qui stocke la série de base ajustée selon les moyennes dans sa cinquième colonne, la série de base ajustée selon QM dans la neuvième colonne, et la série de base originale dans sa troisième colonne; et
  - Example\_Ref1\_F.pdf (ou Example\_F.pdf). Le fichier Example\_Ref1\_F.pdf contient six graphiques: (i) différence entre base et référence séries; (ii) série de base



d'anomalies (c.-à-d. la série de base dont le cycle annuel moyen est soustrait aussi référée à la série de base désaisonnalisée) et ajustement par modèle de régression de multi-phase; (iii) série de base, sauts estimés et tendance linéaire; (iv) série de base ajustée selon les moyennes; (v) série de base ajustée selon QM (les deux dernières ajustées selon le segment de base), et (vi) la distribution des ajustements QM. Le fichier Example\_Ref1\_F.pdf contient tous les graphiques avec l'exception du premier. Veuillez lire les sections 3.2 et 3.3 pour avoir une description du contenu de ces fichiers.

- (9) Si vous voulez analyser les données de température (ou des données gaussiennes en général) et que vous voulez ajuster les températures quotidiennes pour les points de changements significatifs que vous avez identifié dans les séries mensuelles avec une station de référence, cliquer sur le bouton **QMadj\_GaussDLY.wRef** pour choisir le fichier des données des températures quotidiennes (disons Example1\_DLY.dat), et sa série de référence (disons Ref1\_DLT.dat) et la liste de points de changement identifiés (voir la fenêtre ci-dessous),



et cliquer sur le bouton **OK** pour estimer et appliquer les ajustements QM avec la série de référence. Cela va produire les trois fichiers suivants dans le répertoire de sortie : (a) Example1\_Ref1\_DLY.dat\_QMadjDLY\_stat.txt, qui contient les données originales et ajustées dans les 4<sup>ème</sup> et 5<sup>ème</sup> colonne respectivement (les colonnes 6, 7, 8 sont les données quotidiennes ajustées par la moyenne, les ajustements QM, et les ajustements moyens, respectivement; colonnes 4+7=colonne 5; colonnes 4+8=colonne 6); et (c) Example1\_Ref1\_DLY.dat\_QMadjDLY\_plot.pdf, qui contient les cinq graphiques suivants : (i) la série de température quotidienne désaisonnalisée et le fit MPR, (ii) la série originale et le fit MPR, (iii) la série quotidienne ajustée

par les ajustements moyens, (iv) la série quotidienne ajustée par les ajustements QM, and (v) la distribution des ajustements QM pour chaque segment.

Si vous n'avez pas de série de référence, cliquer sur le bouton **QMadj\_GaussDLY** pour choisir le fichier des données des températures quotidiennes (disons Example1\_DLY.dat) et la liste de points de changement identifiés, et cliquer sur le bouton **OK** pour estimer et appliquer les ajustements QM. Cela va produire les fichiers suivants dans le répertoire de sortie :

- (a) Example1\_DLY.dat\_QMadjDLY\_stat.txt, qui contient les résultats graphiques;
- (b) Example1\_DLY.dat\_QMadjDLY\_data.dat, qui contient les données originales et ajustées dans la 4ième et 5ième colonnes respectivement; et
- (c) Example1\_DLY.dat\_QMadjDLY\_plot.pdf, qui contient les cinq graphiques:
  - (i) la série de température quotidienne désaisonnalisée et le fit MPR, (ii) la série originale et le fit MPR, (iii) la série quotidienne ajustée par les ajustements moyens, (iv) la série quotidienne ajustée par les ajustements QM, and (v) la distribution des ajustements QM pour chaque segment. **Il faut exercer de la prudence lorsqu'on ajuste des séries de données sans utiliser de bonnes séries de référence.**

Si vous voulez analyser les données de précipitation et que vous voulez ajuster les **précipitations quotidiennes** pour les points de changement significatifs qui ont été identifiés dans les séries mensuelles correspondantes, veuillez appeler la fonction *StepSize\_dlyPrcp* dans le progiciel RHtests\_dlyPrcp (aussi disponible au même site) pour ajuster les précipitations quotidiennes pour les points de changement identifiés (veuillez vous référer au guide abrégé pour RClmDex et RHtest qui est aussi disponible sur ce site). Veuillez noter que les ajustements QM peuvent être problématiques si la rupture est présente dans la fréquence des précipitations mesurés (voir Wang et al. 2010 pour plus de détails).

En plus des fonctions de l'interface graphique ci-dessus, RHtestsV4 comporte aussi six fonctions pour détecter des sauts (écarts de la moyenne) dans les séries annuelles, mensuelles ou quotidiennes sans interface graphique. On doit d'abord cliquer sur le bouton **Quit**, puis appeler ces fonctions à l'invite R (voir les sections 3.2 et 3.3 pour plus de détails).

### 3.2 Cas « avec série de référence »

Le test utilisé dans ce cas est le test  $t$  maxima avec pénalité (Wang et al. 2007, Wang 2008a), qui suppose que la série temporelle testée a une tendance nulle et des erreurs gaussiennes. Le résultat de la série de base moins la série de référence est testé pour identifier la ou les position(s) et le degré de signification du ou des points de changement, et un modèle de régression multi-phase (MPR) avec une tendance commune est aussi appliqué à la fin aux anomalies (relativement au cycle annuel moyen) de la série de base afin d'obtenir les estimés finaux de l'ordre de grandeur des sauts (voir la section *Appendice* du document Wang 2008a pour plus de détails). Dans MPR, le cycle annuel,



la tendance linéaire et l'autocorrélation de retard de 1 sont estimés utilisant des procédures itératives tout en expliquant tous les sauts identifiés (Wang 2008a)

Dans ce cas, les cinq procédures détaillées sont :

(T1) Appelez la fonction *FindU.wRef* pour identifier tous les points de changement de type 1 de la série de base moins la série de référence en entrant ce qui suit à l'invite R :

```
FindU.wRef(Bseries="C:/inputdata/Bfile.csv", MissingValueCode="-999.0",
           p.lev=0.95, Iadj=10000, Mq=10, Ny4a=0, output="C:/results/OutFile",
           Rseries="C:/inputdata/Rfile.csv")
```

Dans ce cas, *C:/inputdata/* est le répertoire des données et *Bfile.csv* et *Rfile.csv* sont les noms des fichiers contenant les séries de base et de référence, respectivement (on teste ici la série de base moins la série de référence) qui sont dans le répertoire des données; *C:/results/* est un répertoire spécifié par l'utilisateur et *OutFile* est un préfixe sélectionné par l'utilisateur comme nom du fichier afin de stocker les résultats et -999.0 est le code de valeur absente qui est utilisé dans les fichiers de données d'entrée *Bfile.csv* et *Rfile.csv* ; *p.lev* est le niveau nominal pour lequel le test est fait (choisir un des niveaux suivants: 0.75, 0.80, 0.90, 0.95, 0.99, et 0.9999), *Iadj* est une valeur entière correspondant au segment pour lequel la série est ajustée (segment de base), avec *Iadj*=10000 correspondant au dernier segment; *Mq* est le nombre de points (catégories) pour lesquels la fonction cumulative de la distribution empirique est estimée; et *Ny4a* est le nombre maximal d'années de données immédiatement avant ou après le point de rupture qui est utilisé pour estimer le PDF (*Ny4a*=0 pour choisir le segment au complet). On peut donner à *Mq* n'importe quelle valeur entière entre 1 and 100, ou si *Mq*=0 le nombre de points sera automatiquement déterminé par la fonction (la fonction remet *Mq* à 1 si 0 est choisi ou à 100 si un plus grand nombre que 100 est choisi. Les valeurs sont par défaut : *p.lev*=0.95, *Iadj*=10000, *Mq*=12, *Ny4a*=0. Il est à noter que le code de donnée manquante (*MissingValueCode*) entré ici **doit être exactement le même** que celui qui est utilisé dans les données; p. ex., on ne peut pas entrer « -999. » au lieu de « -999.0 » lorsque « -999.0 » est utilisé dans la série de données d'entrée; sinon cela produirait des données erronées. Il est aussi à noter que les chaînes de caractères doivent être insérées entre guillemets doubles, comme ci-dessus. Après un appel réussi, cette fonction produit les cinq fichiers suivants dans le répertoire de sortie:

- *OutFile\_1Cs.txt* (ou *OutFile\_mCs.txt*) : Le premier chiffre de la première ligne de ce fichier est le nombre de points de changement identifiés dans la série testée. Si ce nombre est  $N_c > 0$ , les  $N_c$  lignes suivantes énumèrent les dates et les statistiques de ces  $N_c$  points de changement. Par exemple : le fichier aura le contenu suivant dans le cas de  $N_c = 10$  :

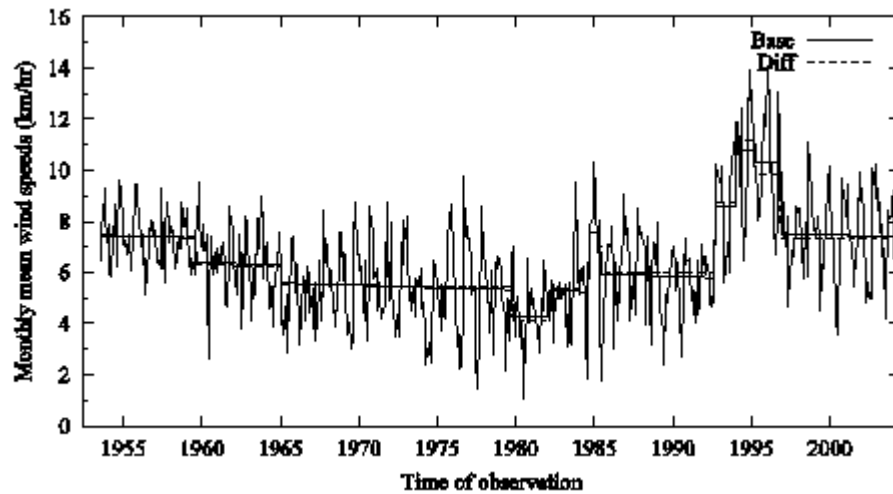
```
10 changepoints in Series InFile.csv
1 Yes 19600200 (1.0000-1.0000) 0.950 5.3553 ( 3.0635- 3.5582)
1 Yes 19650700 (0.9999-0.9999) 0.950 4.1634 ( 3.0977- 3.6009)
```

1	Yes	19800300	(1. 0000–1. 0000)	0. 950	5. 2305	(	3. 0905–	3. 5927)
1	Yes	19820800	(1. 0000–1. 0000)	0. 950	4. 6415	(	3. 0237–	3. 5000)
1	Yes	19850200	(1. 0000–1. 0000)	0. 950	4. 7195	(	3. 0005–	3. 4712)
1	Yes	19851100	(0. 9999–0. 9999)	0. 950	3. 8607	(	3. 0457–	3. 5283)
1	Yes	19930200	(1. 0000–1. 0000)	0. 950	9. 1678	(	3. 0469–	3. 5342)
1	Yes	19940600	(1. 0000–1. 0000)	0. 950	5. 9064	(	2. 9845–	3. 4552)
1	?	19950900	(0. 9965–0. 9971)	0. 950	3. 0340	(	2. 9929–	3. 4612)
1	Yes	19970400	(1. 0000–1. 0000)	0. 950	8. 2999	(	3. 0497–	3. 5398)

La première colonne (les 1) indique que ce sont les points de changement de type 1 (aussi indiqué par le « 1Cs » du nom de fichier). La deuxième colonne indique si le point de changement est significatif ou non pour le type fourni à la première colonne; ils sont tous « Yes » dans ce fichier \*\_1Cs.txt, mais dans les autres fichiers \*\_Cs.txt ils peuvent être : (1) “Yes” (significatif); (2) “No” (pas significatif pour le type fourni dans la première colonne); (3) “?” (peut-être significatif pour le type fourni dans la première colonne, et (4) “YifD” (significatif si documenté, c’est-à-dire supporté par l’historique de la station). La troisième colonne liste les dates des points de changement YYYYMMDD : par exemple 196002100 indique février 1960. Les nombres dans la quatrième colonne fournissent l’intervalle de confiance à 95% de la valeur p, qui est estimée assumant que le point de changement est documenté). La valeur p nominale (niveau de confiance) est fournie dans la cinquième colonne. Les trois dernières colonnes fournissent la valeur du test statistique  $PF_{max}$  et l’intervalle de confiance à 95% des centiles de  $PF_{max}$  qui correspondent au niveau de confiance nominal. Une copie du fichier OutFile\_1Cs.txt est sauvegardée dans le fichier OutFile\_mC.txt pour des modifications postérieures (pour que l’original reste inchangé).

- OutFile\_Ustat.txt : En plus de tous les résultats stockés dans le fichier OutFile\_1Cs.txt, ce fichier de sortie contient les paramètres estimés de la correction par modèle de régression à  $(N_c + 1)$  phase(s), y compris l’ampleur des sauts identifiés, la tendance linéaire et l’autocorrélation avec retard de 1 de la série de base.
- OutFile\_U.dat : Dans ce fichier, la 2<sup>e</sup> colonne contient les dates d’observation; la 3<sup>e</sup> colonne est la série de base originale; la 4<sup>e</sup> et 5<sup>e</sup> colonnes sont la tendance linéaire estimée de la série de base et de la série ajustées selon la moyenne pour les sauts estimés de la série de base moins la série de référence; la 6<sup>e</sup> et 7<sup>e</sup> colonnes sont similaires à la 4<sup>e</sup> et 5<sup>e</sup> colonnes mais pour les sauts estimés de la série de base ajustée pour les saisons; la 8<sup>e</sup> et la 9<sup>e</sup> colonnes sont les série de base ajustées pour les saisons (c.-à-d. la série de base dont le cycle annuel moyen est soustrait) et sa correction par modèle de régression à phases multiples; la 10<sup>e</sup> colonne est cycle annuel moyen estimé avec la tendance linéaire et écarts de la moyenne; la 11<sup>e</sup> colonne est la série de base ajustée selon QM (ici, les ajustements QM sont estimés utilisant une série de référence); et la 12<sup>e</sup> colonne est le model de régression multi-phase de la série de base dessaisonnalisée sans tenir compte des sauts (ignorant tous les sauts identifiés).

- OutFile\_U.pdf : Ce fichier a six graphiques : (i) la série de base moins la série de référence, (ii) la série de base d'anomalies ainsi que sa correction par modèle de régression à phases multiples; (iii) la série de base ainsi que les sauts estimés et la tendance linéaire; (iv) la série de base ajustée selon la moyenne, (v) la série de base ajustée selon QM, et(vi) la distribution des ajustements QM (ils sont estimés utilisant une série de référence) . Voici un exemple du troisième volet graphique :



La tendance en ligne solide représente la correction par le modèle de régression et la ligne en pointillée représente les sauts estimés de la série base moins la série de référence.

Si aucun point de changement significatif n'a été identifié à cette étape, la série temporelle testée peut être déclarée homogène et il n'est pas nécessaire d'appliquer d'autres tests à cette série.

**(T2) Si vous connaissez des points de changement qui peuvent causer des sauts, ajoutez les dates correspondant à ces points au fichier Example\_mC.s.txt s'ils ne sont pas déjà là, et allez à et allez à la procédure (T4) maintenant. Si aucune métadonnée n'est disponible ou si vous voulez détecter seulement les points de changement significatifs même sans le soutien de métadonnées (c.-à-d. les points de changement de type 1), allez aussi à la procédure (T4).** Sinon, appelez la fonction *FindUD.wRef* pour identifier tous les points de changement de type 0 probables de la série de base, en présence de tous les points de changement de type 1 énumérés dans le fichier OutFile\_1Cs.txt, en entrant ce qui suit à l'invite R :

```
FindUD.wRef(Bseries="C:/inputdata/Bfile.csv", MissingValueCode="-999.0",
p.lev=0.95, Iadj=10000, Mq=10, Ny4a=0, InCs="C:/results/OutFile_1Cs.txt",
output="C:/results/OutFile", Rseries="C:/inputdata/Rfile.csv")
```

Dans ce cas, le fichier OutFile\_1Cs.txt contient tous les points de changement de type 1 identifiés en appelant *FindU.wRef* en (T1) ci-dessus et tous les autres fichiers

sont les mêmes qu'en (T1). Après un appel réussi, cette fonction produit les cinq fichiers suivants : OutFile\_pCs.txt et OutFile\_mCs.txt, OutFile\_UDstat.txt, OutFile\_UD.pdf et OutFile\_UD.dat. Le contenu de ces fichiers est semblable à celui des fichiers pertinents de (T1), sauf que les points de changement qui sont maintenant modélisés sont ceux qui sont énumérés dans le fichier OutFile\_pCs.txt ou OutFile\_mCs.txt, qui contiennent tous les points de changement de type 1 énumérés dans OutFile\_1Cs.txt, **plus** tous les points de changement de type 0 probables. Le fichier OutFile\_mCs.txt est maintenant une copie de OutFile\_pCs.txt, ce qui permettra des modifications ultérieures. Pour l'exemple ci-dessus, ce fichier contient maintenant 18 points de changement de type 0 (pas présentés ici), en plus des dix points de changement de type 1.

Puisque cette étape peut aider à resserrer l'investigation des métadonnées, il est suffisant d'appliquer ces fonctions sur des séries mensuelles ou annuelles, et il n'est pas nécessaire, et cela prendra trop de temps, de les appliquer sur les données quotidiennes. Lorsqu'on analyse les séries des données quotidiennes, on devrait appliquer des fonctions sur les séries mensuelles correspondantes.

**(T3)** Comme il a été mentionné précédemment, les points de changement de type 0 **pourraient être** significatifs statistiquement au niveau préétabli de signification **seulement si** ils sont soutenus par des métadonnées fiables. De plus, certains des points de changement de type 1 identifiés pourraient aussi être soutenus par des métadonnées et les dates de changement exactes pourraient différer légèrement des dates qui ont été identifiées statistiquement. On devrait donc alors se concentrer sur les métadonnées disponibles, particulièrement sur les dates de tous les points de changement (type 1 ou type 0) énumérés dans OutFile\_mCs.txt. **Conservez seulement les points de changement de type 0 qui sont soutenus par des métadonnées, ainsi que tous les points de changement de type 1.** Modifiez les dates de points de changement identifiés statistiquement aux dates documentées de changement (obtenues au moyen de métadonnées hautement fiables) au besoin. Pour l'exemple ci-dessus, il y a seulement deux des 18 points de changement de type 0 qui sont soutenus par des métadonnées et les dates exactes de ces décalages sont en Janvier 1974 et en Novembre 1975. Dans ce cas, on devrait modifier le fichier OutFile\_mCs.txt comme suit :

	12	changepoints	in Series	InFile.csv							
1	Yes	19600200	(	1.0000-	1.0000)	0.950	5.3553	(	2.9659-	3.4415)	
1	Yes	19650700	(	0.9998-	0.9999)	0.950	3.9531	(	2.9775-	3.4575)	
0	Yes	19740200	(	0.9999-	1.0000)	0.950	4.1114	(	2.9599-	3.4299)	
0	Yes	<b>19751100</b>	<b>19760100</b>	(	1.0000-	1.0000)	0.950	4.5880	(	2.9330-	3.4003)
1	Yes	19800300	(	1.0000-	1.0000)	0.950	6.1548	(	2.9403-	3.4003)	
1	Yes	19820800	(	1.0000-	1.0000)	0.950	4.6415	(	2.9283-	3.3883)	
1	Yes	19850200	(	1.0000-	1.0000)	0.950	4.7195	(	2.9083-	3.3588)	
1	Yes	19851100	(	0.9999-	0.9999)	0.950	3.8607	(	2.9490-	3.4163)	
1	Yes	19930200	(	1.0000-	1.0000)	0.950	9.1678	(	2.9515-	3.4215)	
1	Yes	19940600	(	1.0000-	1.0000)	0.950	5.9064	(	2.8923-	3.3428)	
1	?	19950900	(	0.9963-	0.9970)	0.950	3.0340	(	2.8983-	3.3503)	
1	Yes	19970400	(	1.0000-	1.0000)	0.950	8.2999	(	2.9543-	3.4243)	

[**Veillez ne pas changer le format des trois premières colonnes**, elles doivent être lues plus tard comme entrants avec un format équivalent au **format (i1,a4,i10)** en FORTRAN.].

Il pourrait aussi arriver qu'il ne soit pas nécessaire d'apporter de modification au fichier OutFile\_mCs.txt (ni pour les valeurs, ni pour les dates des points de changement; et les fichiers OutFile\_pCs.txt et OutFile\_mCs.txt demeurent donc identiques); dans ce cas la procédure (T4) ci-dessous peut être omise.

(T4) Appelez la fonction *StepSize.wRef* pour estimer à nouveau le degré de signification et l'ordre de grandeur des points de changement énumérés dans OutFile\_mCs.txt :

```
StepSize.wRef(Bseries="C:/inputdata/Bfile.csv", MissingValueCode="-999.0",
              p.lev=0.95, Iadj=10000, Mq=10, Ny4a=0,
              InCs="C:/results/OutFile_mCs.txt", output="C:/results/OutFile",
              Rseries="C:/inputdata/Rfile.csv")
```

ce qui produira les cinq fichiers suivants dans le répertoire de sortie:

- OutFile\_fCs.txt (ou OutFile\_mCs.txt), semblable au fichier d'entrée OutFile\_mCs.txt, sauf qu'il contient de nouveaux estimés du degré de signification/des valeurs des points de changement énumérés dans le fichier OutFile\_mCs.txt. Le résultat ressemble à ceci :

```

12 changepoints in Series InFile.csv
1 Yes 19600200 ( 1.0000- 1.0000) 0.950 5.3553 ( 2.9659- 3.4415)
1 Yes 19650700 ( 0.9998- 0.9999) 0.950 3.9531 ( 2.9775- 3.4575)
0 Yes 19740200 ( 0.9999- 1.0000) 0.950 4.1114 ( 2.9599- 3.4299)
0 Yes 19751100 ( 1.0000- 1.0000) 0.950 4.5880 ( 2.9330- 3.4003)
1 Yes 19800300 ( 1.0000- 1.0000) 0.950 6.1548 ( 2.9403- 3.4003)
1 Yes 19820800 ( 1.0000- 1.0000) 0.950 4.6415 ( 2.9283- 3.3883)
1 Yes 19850200 ( 1.0000- 1.0000) 0.950 4.7195 ( 2.9083- 3.3588)
1 Yes 19851100 ( 0.9999- 0.9999) 0.950 3.8607 ( 2.9490- 3.4163)
1 Yes 19930200 ( 1.0000- 1.0000) 0.950 9.1678 ( 2.9515- 3.4215)
1 Yes 19940600 ( 1.0000- 1.0000) 0.950 5.9064 ( 2.8923- 3.3428)
1 ? 19950900 ( 0.9963- 0.9970) 0.950 3.0340 ( 2.8983- 3.3503)
1 Yes 19970400 ( 1.0000- 1.0000) 0.950 8.2999 ( 2.9543- 3.4243)
```

Une copie de OutFile\_fCs.txt est aussi stockée comme fichier OutFile\_mCs.txt file (que la version d'entrée est mise à jour avec les nouveaux estimés de signification et de statistiques) pour analyse ultérieure.

- OutFile\_Fstat.txt, semblable au fichier OutFile\_Ustat.txt ou OutFile\_UDstat.txt ci-dessus, sauf que les points de changement dont il est tenu compte ici sont ceux énumérés dans le fichier OutFile\_mCs.txt.
- OutFile\_F.dat. semblable au fichier OutFile\_U.dat ou OutFile\_UD.dat ci-dessus, sauf que les points de changement dont il est tenu compte ici sont ceux énumérés dans le fichier OutFile\_mCs.txt.

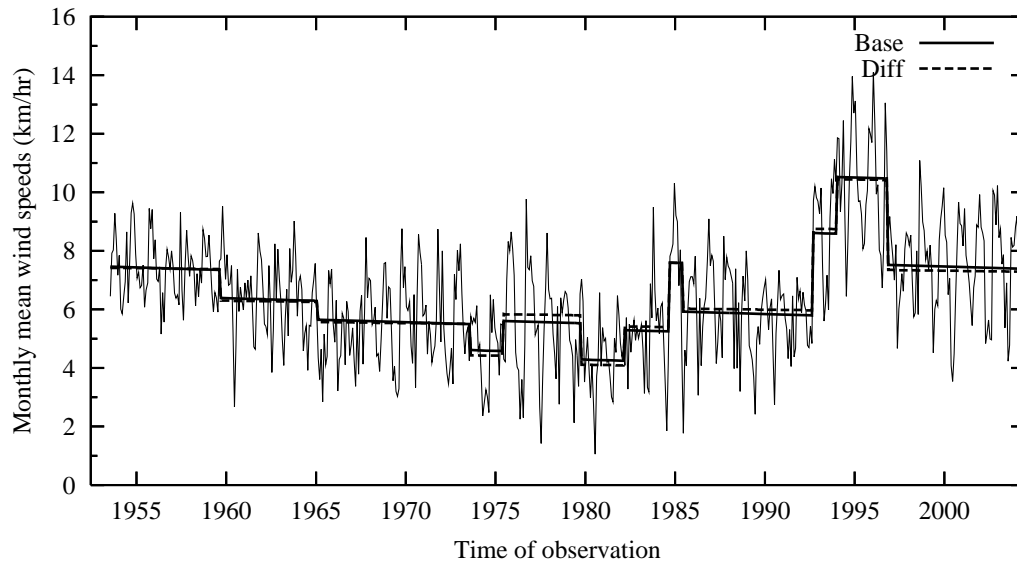
- OutFile\_F.pdf. semblable au fichier OutFile\_U.pdf ou OutFile\_UD.pdf ci-dessus, sauf que les points de changement dont il est tenu compte ici sont ceux énumérés dans le fichier OutFile\_mCs.txt.

(T5) À ce point, on doit analyser les résultats pour déterminer si le plus petit saut parmi tous les sauts/points de changement demeure significatif (l'ordre de grandeur des sauts est inclus dans le fichier OutFile\_Fstat.txt ou OutFile\_Ustat.txt file). Pour cela, on doit comparer la valeur  $p$  (s'il est de type 0) ou la valeur  $PT_{max}$  (s'il est de type 1) du plus petit saut avec la plage d'incertitude correspondante à 95 %. On peut estimer que le plus petit saut est significatif si sa valeur  $p$   $PT_{max}$  est plus grande que la limite supérieure correspondante et n'est pas significatif s'il est plus petit que la limite inférieure. Toutefois, si la valeur  $p$  ou la valeur  $PT_{max}$  se trouve dans la plage d'incertitude correspondante à 95 %, on **doit déterminer subjectivement** si on doit considérer que ce point de changement est significatif ou non (la consultation du graphique dans OutFile\_F.pdf ou OutFile\_U.pdf pourrait aider); cela est dû à l'incertitude inhérente à l'estimé de la valeur inconnue de l'autocorrélation avec retard de 1 de la série (voir Wang 2008a).

Si le plus petit saut n'est pas significatif, supprimez-le du fichier OutFile\_mCs.txt et appelez à nouveau la fonction *StepSize.wRef* avec la nouvelle liste modifiée de points de changement. Pour l'exemple ci-dessus, le point de changement de septembre 1995 a été déterminé pas signifiant à cause du manque de support des métadonnées et il peut être ou non un point de changement de type 1 (puisque  $PT_{max}$  est dans la plage d'incertitude correspondante à 96% des percentiles 95). On le supprime donc du fichier OutFile\_mCs.txt et on appelle à nouveau la fonction *StepSize.wRef*, ce qui produira donc le fichier OutFile\_mCs.txt mis à jour :

11 changepoints in Series InFile.csv							
1	Yes	19600200	(	1.0000-	1.0000)	0.950	5.3553 ( 3.0020- 3.4871)
1	Yes	19650700	(	0.9998-	0.9999)	0.950	3.9531 ( 3.0136- 3.5017)
0	Yes	19740200	(	0.9999-	1.0000)	0.950	4.1114 ( 2.9960- 3.4771)
0	Yes	19751100	(	1.0000-	1.0000)	0.950	4.5880 ( 2.9703- 3.4445)
1	Yes	19800300	(	1.0000-	1.0000)	0.950	6.1548 ( 2.9764- 3.4476)
1	Yes	19820800	(	1.0000-	1.0000)	0.950	4.6415 ( 2.9644- 3.4325)
1	Yes	19850200	(	1.0000-	1.0000)	0.950	4.7195 ( 2.9444- 3.4025)
1	Yes	19851100	(	0.9999-	0.9999)	0.950	3.8607 ( 2.9863- 3.4605)
1	Yes	19930200	(	1.0000-	1.0000)	0.950	9.1678 ( 2.9876- 3.4661)
1	Yes	19940600	(	0.9999-	1.0000)	0.950	4.3806 ( 2.9564- 3.4245)
1	Yes	19970400	(	1.0000-	1.0000)	0.950	12.6841 ( 2.9964- 3.4776)

On doit recommencer cette procédure d'estimation, c.-à-d. répéter l'appel de la fonction *StepSize.wRef*, jusqu'à ce que **tous** les points de changement énumérés dans OutFile\_fCs.txt ou OutFile\_mCs.txt soient estimés comme étant significatifs. Pour l'exemple ci-dessus, tous les 11 points de changements sont significatifs de type 1 (ils sont signifiant même sans support des métadonnées); le graphique final est donc :



Notez que deux possibilités d'estimés de saut (ajustements) sont fournis ici : un premier estimé à partir de la série base et un deuxième à partir de la série de différence. Les usagers ont le choix et ils sont responsables pour déterminer quels ajustements à faire. L'attention sera exercée quand il y a une anomalie significative entre les deux possibilités d'estimés de sauts, qui pourraient résulter ou de l'inhomogénéité de la série de référence, ou de l'existence de quelques autres points de changements qui ne sont pas identifiés dans la série de base (la chance pour un tel échec est égale au niveau  $\alpha$ ). En cas d'inhomogénéité de série de référence, les estimés de sauts de la série de différence ne devraient pas être utilisés (autrement elle présenterait un ou plusieurs nouveaux points de changements à la série base); les ajustements estimés à partir de la série base sont meilleurs dans ce cas-ci et devraient être utilisés. Quand tous les points de changements significatifs sont expliqués et la série de référence est homogène, les deux possibilités d'estimés de sauts devraient être semblables, comme montré dans l'exemple ci-dessus.

### 3.3 Cas « sans série de référence »

**Avertissement** : Les résultats de la détection des points de changements sans l'utilisation d'une station de référence sont moins sûrs et ont besoin plus d'analyse. On ne devrait pas faire l'homogénéisation de données d'une façon automatique sans une série de référence (c'est-à-dire sans plus d'analyse manuelle des résultats statistiques). Il faut exercer de la prudence lorsqu'on ajuste des séries de données sans utiliser de bonnes séries de référence.

Le test appliqué dans cette section est le test de  $F$  maxima avec pénalité (Wang 2008a et 2008b), qui permet que la série temporelle testée ait une tendance linéaire pendant toute la période de l'enregistrement de données (c.-à-d. qu'il n'y a pas de variation de la composante tendance; voir Wang 2003) avec le cycle annuel, l'autocorrélation de retard de 1, et la tendance linéaire de la série de base estimés tout en expliquant tous les sauts identifiés (Wang 2008a). Cela correspond fondamentalement au

même cas que dans la section 3.1 ci-dessus, sauf qu'il n'y a pas d'interface graphique. La série temporelle qui est testée peut être une série de base (on a alors un vrai cas sans série de référence) ou une série de base moins la série de référence (dans une seule série prête à traiter). Ce dernier cas est recommandé **seulement si** on soupçonne une différence entre les tendances de la série de base et la série de référence; dans ce cas, les estimés des paramètres pour la série de base doivent être obtenus en effectuant un ou des appels additionnels à la fonction *StepSize*, avec la série de base et les points de changement qui sont identifiés à partir de la série de base moins la série de référence [voir la procédure (F5) de la présente section].

Dans ce cas, les cinq procédures détaillées sont les suivantes :

(F1) Appelez la fonction *FindU* pour identifier tous les points de changement de type 1 de la série d'entrée (InSeries) en entrant ce qui suit à l'invite R :

```
FindU(InSeries="C:/inputdata/InFile.csv", MissingValueCode="-999.0",
      p.lev=0.95, Iadj=10000, Mq=10, Ny4a=0, output="C:/results/OutFile")
```

Dans ce cas, C:/inputdata/ est le répertoire de données et InFile.csv est le nom du fichier contenant la série de données à tester; tandis que C:/results/ est un répertoire spécifié par l'utilisateur and OutFile est un préfixe sélectionné par l'utilisateur comme nom du fichier afin de stocker les résultats et -999.0 est le code de valeur absente utilisé dans le fichier d'entrée de données InFile.csv ; *p.lev* est le niveau nominal pour lequel le test est fait (choisir un des niveaux suivants: 0.75, 0.80, 0.90, 0.95, 0.99, et 0.9999), ***Iadj* est une valeur entière correspondant au segment pour lequel la série est ajustée (segment de base), avec *Iadj*=10000 correspondant au dernier segment; *Mq* est le nombre de points (catégories) pour lesquels la fonction cumulative de la distribution empirique est estimée; et *Ny4a* est le nombre maximal d'années de données immédiatement avant ou après le point de rupture qui est utilisé pour estimer le PDF (*Ny4a*=0 pour choisir le segment au complet). On peut donner à *Mq* n'importe quelle valeur entière entre 1 and 100, ou si *Mq*=0 le nombre de points sera automatiquement déterminé par la fonction (la fonction remet *Mq* à 1 si 0 est choisi ou à 100 si un plus grand nombre que 100 est choisi. Les valeurs sont par défaut : *p.lev*=0.95, *Iadj*=10000, *Mq*=12, *Ny4a*=0.** Il est à noter que le code de valeur absente (MissingValueCode) entré ici **doit être exactement le même** qui est utilisé dans les données; p. ex., on ne peut pas entrer « -999. » au lieu de « -999.0 » lorsque « -999.0 » est utilisé dans la série de données d'entrée; sinon cela produira des données erronées. Il est aussi à noter que les chaînes de caractères doivent être insérées entre guillemets doubles, comme ci-dessus. Après un appel réussi, cette fonction produit les cinq fichiers suivants dans le répertoire de sortie:

- OutFile\_1Cs.txt (et OutFile\_mCs.txt) : Le premier chiffre de la première ligne de ce fichier est le nombre de points de changement identifiés dans la série testée. Si ce nombre est  $N_c > 0$ , les  $N_c$  lignes suivantes énumèrent les dates et les statistiques de ces  $N_c$  points de changement. Par exemple : le fichier aura le contenu suivant dans le cas de  $N_c = 3$  :



```

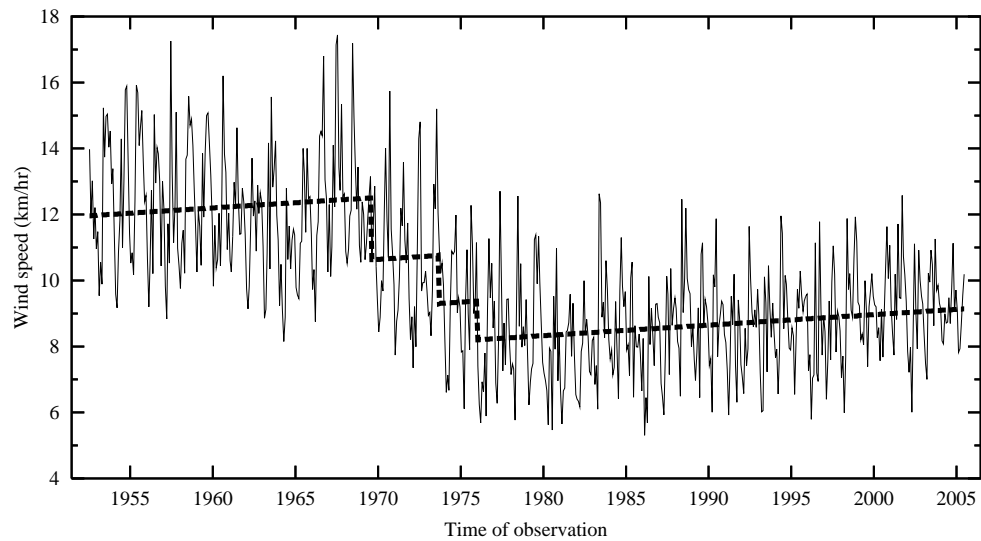
3 changepoints in Series InFile.csv
1 Yes 19700100 (1.0000-1.0000) 0.950 49.2060 ( 14.2857- 19.2003)
1 Yes 19740200 (1.0000-1.0000) 0.950 23.3712 ( 13.2758- 17.7289)
1 Yes 19760600 (1.0000-1.0000) 0.950 21.4244 ( 14.4950- 19.5295)

```

La première colonne (les 1) indique que ce sont les points de changement de type 1 (aussi indiqué par le « 1Cs » du nom de fichier). La deuxième colonne indique si le point de changement est significatif ou non pour le type fourni à la première colonne; ils sont tous « Yes » dans ce fichier \*\_1Cs.txt, mais dans les autres fichiers \*\_Cs.txt ils peuvent être : (1) “Yes” (significatif); (2) “No” (pas significatif pour le type fourni dans la première colonne); (3) “?” (peut-être significatif pour le type fourni dans la première colonne, et (4) “YifD” (significatif si documenté, c’est-à-dire supporté par l’historique de la station). La troisième colonne liste les dates des points de changement YYYYMMDD : par exemple 19700100 indique janvier 1970. Les nombres dans la quatrième colonne (entre parenthèses) fournissent l’intervalle de confiance à 95% de la valeur p, qui est estimée assumant que le point de changement est documenté (ainsi cette valeur est très élevée pour un point de changement significatif de type 1). La valeur p nominale (niveau de confiance) est fournie dans la cinquième colonne. Les trois dernières colonnes fournissent la valeur du test statistique  $PF_{\max}$  et l’intervalle de confiance à 95% des centiles de  $PF_{\max}$  qui correspondent au niveau de confiance nominal. Une copie du fichier OutFile\_1Cs.txt est sauvegardée dans le fichier OutFile\_mCs.txt pour des modifications postérieures (pour que l’original reste inchangé).

- OutFile\_Ustat.txt : En plus de tous les résultats stockés dans le fichier OutFile\_1Cs.txt, ce fichier de sortie contient les paramètres estimés de la correction par modèle de régression à  $(N_c + 1)$  phase(s), y compris la taille des sauts identifiés, la tendance linéaire et l’autocorrélation avec retard de 1 de la série testée.
- OutFile\_U.dat : Ce fichier contient les dates d’observation (2<sup>e</sup> colonne), la série de base originale (3<sup>e</sup> colonne), la tendance linéaire estimée et les écarts de la moyenne de la série de base (4<sup>e</sup> colonne), la série de base ajustée selon la moyenne (5<sup>e</sup> colonne), la série de base d’anomalies (c.-à-d. la série de base dont le cycle annuel moyen est soustrait) et sa correction par modèle de régression à phases multiples (6<sup>e</sup> et 7<sup>e</sup> colonnes, respectivement) et cycle annuel moyen estimé avec la tendance linéaire et écarts de la moyenne (8<sup>e</sup> colonne), les ajustements QM, (9<sup>e</sup> colonne), le modèle de régression de multi-phase pour la série de base désaisonnalisée sans compté les sauts identifiés (10<sup>e</sup> colonne).
- OutFile\_U.pdf : Ce fichier a cinq graphiques : (i) la série de base d’anomalies (c.-à-d. anomalies relatives au cycle annuel moyen de la série de base) ainsi que sa correction par modèle de régression à phases multiples; (ii) la série de base ainsi que les sauts estimés de la moyenne et la tendance linéaire; (iii) série

de base ajustée selon les moyennes, (iv) série de base ajustée selon QM (les deux dernières ajustées selon le segment de base) et (v) la distribution des ajustements QM. Voici un exemple du second volet graphique :



S'il n'y a pas de point de changement significatif identifié, la série temporelle testée peut être déclarée homogène et il n'est pas nécessaire d'appliquer d'autres tests à cette série.

**(F2) Si vous connaissez des points de changement qui peuvent causer des sauts, ajoutez les dates correspondant à ces points au fichier Example\_mCs.txt s'ils ne sont pas déjà là, et allez à (F4) maintenant. Si aucune métadonnée n'est disponible ou si vous voulez détecter seulement les points de changement significatifs même sans le soutien de métadonnées (c.-à-d. les points de changement de type 1), aussi allez à (F4).** Sinon, appelez la fonction *FindUD* pour identifier tous les points de changement de type 0 de la série, en présence de tous les points de changement de type 1 énumérés dans le fichier OutFile\_1Cs.txt, en entrant ce qui suit à l'invite R :

```
FindUD(InSeries="C:/inputdata/InFile.csv", MissingValueCode="-999.0",
      p.lev=0.95, Iadj=10000, Mq=10, Ny4a=0, InCs="C:/results/OutFile_1Cs.txt",
      output="C:/results/OutFile")
```

Dans ce cas, le fichier OutFile\_1Cs.txt contient tous les points de changement de type 1 identifiés en appelant *FindU* en(F1) ci-dessus et tous les autres fichiers sont les mêmes qu'en (F1). Dans ce cas, un appel réussi produit aussi cinq fichiers : OutFile\_pCs.txt et OutFile\_mCs.txt, OutFile\_UDstat.txt, OutFile\_UD.pdf et OutFile\_UD.dat. Le contenu de ces fichiers est semblable à celui des fichiers pertinents de (F1), sauf que les points de changement qui sont maintenant modélisés sont ceux qui sont énumérés dans le fichier OutFile\_pCs.txt ou OutFile\_mCs.txt, qui contiennent tous les points de changement de type 1 énumérés dans OutFile\_1Cs.txt, **plus** tous les points

de changement de type 0 probables. Le fichier OutFile\_mCs.txt est maintenant une copie de OutFile\_pCs.txt, ce qui permettra des modifications ultérieures.

Puisque cette étape peut aider à resserrer l'investigation des métadonnées, il est suffisant d'appliquer ces fonctions sur des séries mensuelles ou annuelles, et il n'est pas nécessaire, et cela prendra trop de temps, de les appliquer sur les données quotidiennes. Lorsqu'on analyse les séries des données quotidiennes, on devrait appliquer des fonctions sur les séries mensuelles correspondantes.

**(F3)** Comme il a été mentionné précédemment, les points de changement de type 0 **pourraient être** significatifs statistiquement au niveau préétabli de signification **seulement si** ils sont soutenus par des métadonnées fiables. De plus, certains des points de changement de type 1 identifiés pourraient être soutenus par des métadonnées et les dates de changement exactes pourraient différer légèrement des dates qui ont été identifiées statistiquement. On devrait donc alors se concentrer sur les métadonnées disponibles, particulièrement sur les dates de tous les points de changement (type 1 ou type 0) énumérés dans le fichier OutFile\_mCs.txt. **Conservez seulement les points de changement de type 0 qui sont soutenus par des métadonnées, ainsi que tous les points de changement de type 1.** Modifiez les dates de points de changement identifiés statistiquement aux dates documentées de changement (obtenues au moyen de métadonnées hautement fiables) au besoin. Par exemple : le fichier original OutFile\_mCs.txt est le suivant :

```
5 changepoints in Series InFile.csv
0 YifD 19661100 (0.9585-0.9631) 0.950 4.5890 ( 13.9334- 18.7098)
1 Yes 19700100 (1.0000-1.0000) 0.950 49.2842 ( 13.2249- 17.6775)
1 Yes 19740200 (1.0000-1.0000) 0.950 23.4044 ( 13.1030- 17.4929)
1 No 19760600 (0.9947-0.9949) 0.950 8.2628 ( 13.0731- 17.4490)
0 YifD 19800500 (0.9611-0.9725) 0.950 5.5643 ( 14.2667- 19.2069)
```

Si on détermine après avoir examiné les métadonnées qu'il existe des causes documentées pour trois sauts et que les dates exactes de ces sauts sont novembre 1966, juillet 1976 et mars 1980, on devrait modifier le fichier OutFile\_mCs.txt comme suit (les chiffres modifiés sont en gras) :

```
5 changepoints in Series InFile.csv
0 YifD 19661100 (0.9585-0.9631) 0.950 4.5890 ( 13.9334- 18.7098)
1 Yes 19700100 (1.0000-1.0000) 0.950 49.2842 ( 13.2249- 17.6775)
1 Yes 19740200 (1.0000-1.0000) 0.950 23.4044 ( 13.1030- 17.4929)
0 19760700 19760600 (0.9947-0.9949) 0.950 8.2628 ( 13.0731- 17.4490)
0 YifD 19800300 19800500 (0.9611-0.9725) 0.950 5.5643 ( 14.2667- 19.2069)
```

**[Veuillez ne pas changer le format des trois premières colonnes :** elles doivent être lues plus tard comme entrants avec un format équivalent au **format (i1, a4, i10)** en FORTRAN.]

Il est à noter que l'exemple ci-dessus est un cas où tous les points de changement de type 0 ont un soutien de métadonnées. Toutefois, il pourrait arriver que le soutien des métadonnées n'est pas trouvé pour certains des points de changement de type 0

identifiés; dans ce cas, tous les points de changement de type 0 sans soutien doivent être supprimés de la liste (voir l'exemple de la section 3.3 ci-dessous). Il pourrait aussi arriver qu'il ne soit pas nécessaire d'apporter de modification au fichier `OutFile_mCs.txt` (ni pour les valeurs, ni pour les dates des points de changement; les fichiers `OutFile_pCs.txt` et `OutFile_mCs.txt` sont donc encore identiques); dans ce cas la procédure (F4) ci-dessous peut être omise.

**(F4)** Appelez la fonction `StepSize` pour estimer à nouveau le degré de signification et l'ordre de grandeur des points de changement énumérés dans `OutFile_mCs.txt`, p. ex., entrez à l'invite R ce qui suit :

```
StepSize(InSeries="C:/inputdata/InFile.csv", MissingValueCode="-999.0",
         p.lev=0.95, Iadj=10000, Mq=10, Ny4a=0,
         InCs="C:/results/OutFile_mCs.txt", output="C:/results/OutFile")
```

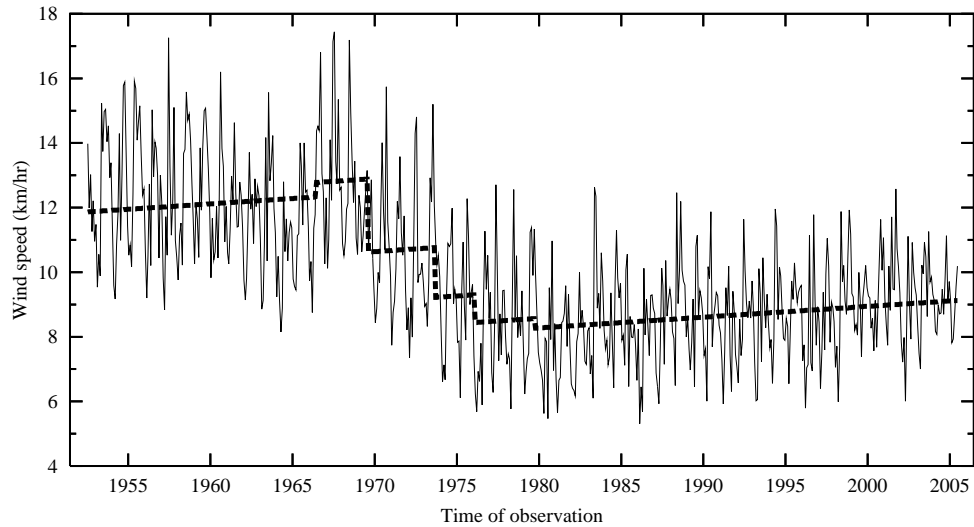
ce qui produira les cinq fichiers suivants dans le répertoire de sortie:

- `OutFile_fCs.txt`, semblable au fichier d'entrée `OutFile_mCs.txt` ci-dessus, sauf qu'il contient de nouveaux estimés du degré de signification/des valeurs des points de changement énumérés dans le fichier d'entrée `OutFile_mCs.txt`. Le résultat ressemble à ceci :

```
5 changepoints in Series InFile.csv
0 YifD 19661100 ( 0.9593- 0.9642) 0.950 4.6639 ( 13.9432- 18.7232)
1 Yes 19700100 ( 1.0000- 1.0000) 0.950 49.2263 ( 13.2341- 17.6898)
1 Yes 19740200 ( 1.0000- 1.0000) 0.950 24.9972 ( 13.1271- 17.5270)
0 YifD 19760700 ( 0.9898- 0.9915) 0.950 7.4609 ( 13.0522- 17.4171)
0 YifD 19800300 ( 0.9586- 0.9648) 0.950 4.8066 ( 14.2757- 19.2186)
```

Une copie de `OutFile_fCs.txt` est aussi stockée comme fichier `OutFile_mCs.txt` (c.-à-d. que la version d'entrée est mise à jour avec les nouveaux estimés de signification et de statistiques) pour analyse ultérieure.

- `OutFile_Fstat.txt`, semblable au fichier `OutFile_Ustat.txt` ou `OutFile_Udstat.txt` ci-dessus, sauf que les points de changement dont il est tenu compte ici sont ceux énumérés dans le fichier `OutFile_mCs.txt`.
- `OutFile_F.dat`, semblable au fichier `OutFile_U.dat` ou `OutFile_UD.dat` ci-dessus, sauf que les points de changement dont il est tenu compte ici sont ceux énumérés dans le fichier `OutFile_mCs.txt`.
- `OutFile_F.pdf`, semblable au fichier `OutFile_U.pdf` ou `OutFile_UD.pdf` ci-dessus, sauf que les points de changement dont il est tenu compte ici sont ceux énumérés dans le fichier `OutFile_mCs.txt`. Pour l'exemple ci-dessus, cela prend l'allure suivante :



(F5) À ce point, on doit analyser les résultats pour déterminer si le plus petit saut parmi tous les sauts/points de changement demeure significatif (l'ordre de grandeur des sauts est inclus dans le fichier `OutFile_Fstat.txt` ou `OutFile_Ustat.txt`). Pour cela, on doit comparer la valeur  $p$  (s'il est de type 0) ou la valeur  $PF_{\max}$  (s'il est de type 1) du plus petit saut avec la plage d'incertitude correspondante à 95 %. On peut estimer que le plus petit saut est significatif si sa valeur  $p$  ou la valeur de  $PF_{\max}$  est plus grande que la limite supérieure correspondante et n'est pas significatif s'il est plus petit que la limite inférieure. Toutefois, si la valeur  $p$  ou la valeur  $PF_{\max}$  se trouve dans la plage d'incertitude correspondante à 95 %, on **doit déterminer subjectivement** si on doit considérer que ce point de changement est significatif ou non (la consultation du graphique dans `OutFile_F.pdf` ou `OutFile_U.pdf` pourrait aider); cela est dû à l'incertitude inhérente à l'estimé de la valeur inconnue de l'autocorrélation avec retard de 1 de la série (voir Wang 2008a).

Si le plus petit saut n'est pas significatif (par exemple : le dernier point de changement ci-dessus n'est pas significatif), supprimez-le du fichier `OutFile_mCs.txt` et appelez à nouveau la fonction `StepSize` avec la nouvelle liste modifiée de points de changement, p. ex., avec cette liste :

```

4 changepoints in Series InFile.csv
0 YifD 19661100 ( 0.9698- 0.9721) 0.950 5.0424 ( 13.9832- 18.7773)
1 Yes 19700100 ( 1.0000- 1.0000) 0.950 48.8569 ( 13.2715- 17.7404)
1 Yes 19740200 ( 1.0000- 1.0000) 0.950 24.5542 ( 13.1641- 17.5771)
0 Yes 19760700 ( 1.0000- 1.0000) 0.950 20.7045 ( 14.3520- 19.3331)

```

On devra recommencer cette procédure d'évaluation (c.-à-d. répéter l'appel de la fonction `StepSize`) jusqu'à ce que **tous** les points de changement énumérés dans `OutFile_fCs.txt` ou `OutFile_mCs.txt` soient estimés significatifs. Par exemple : si le premier point de changement ci-dessus (maintenant le plus petit décalage parmi les quatre) est aussi déterminé non significatif, on devrait le supprimer et appeler à

nouveau la fonction *StepSize* avec les trois points de changement restant, ce qui produira les nouveaux estimés suivants dans le fichier OutFile\_fCs.txt :

```

3 changepoints in Series InFile.csv
1 Yes 19700100 ( 1.0000- 1.0000) 0.950 49.1132 ( 14.2778- 19.1894)
1 Yes 19740200 ( 1.0000- 1.0000) 0.950 25.0495 ( 13.2838- 17.7413)
0 Yes 19760700 ( 1.0000- 1.0000) 0.950 21.0748 ( 14.4869- 19.5183)

```

Dans ce cas, les trois points de changement sont significatifs même sans le soutien de métadonnées, car chacune des valeurs correspondantes de  $PF_{\max}$  (colonne 5 ci-dessus) est supérieure à la limite supérieure de son percentile au niveau nominal choisi (le dernier nombre de chaque ligne). Par conséquent, les résultats obtenus au moyen du dernier appel de la fonction *StepSize* sont les résultats finals pour la série testée.

Il est à noter que si la série testée ci-dessus est une série de base moins la série de référence (et non la série de base elle-même), on doit recommencer d'appeler à nouveau la fonction *StepSize*, avec la série de base comme fichier InFile.csv et les points de changement énumérés dans la plus récente fonction de OutFile\_fCs.txt ou OutFile\_mCs.txt, pour obtenir les estimés finaux de paramètres pour la série de base.

## Références

- Vincent, L. A., X. L. Wang, E. J. Milewska, H. Wan, Y. Feng, and V. Swail, 2012: A Second Generation of Homogenized Canadian Monthly Surface Air Temperature for Climate Trend Analysis, *JGR-Atmospheres*, 117, D18110, doi:10.1029/2012JD017859.
- Wang, X. L., H. Chen, Y. Wu, Y. Feng, and Q. Pu, 2010: New techniques for detection and adjustment of shifts in daily precipitation data series. *J. Appl. Meteor. Climatol.*, **49** (No. 12), 2416-2436. DOI: 10.1175/2010JAMC2376.1
- Wang, X. L., 2008a: Accounting for autocorrelation in detecting mean-shifts in climate data series using the penalized maximal  $t$  or  $F$  test. *J. Appl. Meteor. Climatol.*, **47**, 2423–2444.
- Wang, X. L., 2008b: Penalized maximal F-test for detecting undocumented mean-shifts without trend-change. *J. Atmos. Oceanic Tech.*, **25** (No. 3), 368-384. DOI:10.1175/2007/JTECHA982.1.
- Wang, X. L., Q. H. Wen, and Y. Wu, 2007: Penalized maximal  $t$  test for detecting undocumented mean change in climate data series. *J. Appl. Meteor. Climatol.*, 46 (No. 6), 916-931. DOI:10.1175/JAM2504.1
- Wang, X. L., 2003: Comments on “Detection of Undocumented Change-points: A Revision of the Two-Phase Regression Model”. *J. Climate*, **16**, 3383-3385.

## Remerciements

Les auteurs souhaitent remercier Xuebin Zhang et Enric Aguilar de leurs commentaires utiles sur la version précédente de ce manuel et Hui Wan, Lucie Vincent and Enric Aguilar de leur aide pour effectuer les essais de ce logiciel. Nous remercions aussi Jeff Robel du National Climatic Data Center de NOAA (USA) pour sa révision et l'édition du manuel de la version précédente (RHtestV2 User Guide). Les auteurs veulent aussi remercier le Service intermédiaire de traduction et de révision d'Environnement Canada pour la traduction de ce rapport et Lucie Vincent pour vérifier la traduction française et les mises à jour.