

DÉTECTION DES TENDANCES ET
DÉPASSEMENT DE NORMES
EN QUALITÉ DE L'EAU

RAPPORT MÉTHODOLOGIQUE DÉTAILLÉ

INRS-Eau
C.P. 7500, Sainte-Foy
Québec, Canada
G1V 4C7

Version 2
Juillet 1988

par

Daniel CLUIS
Claude LABERGE
Caroline HOULE

pour

Service Protection et Conservation
Environnement Canada - Région du Québec

Rapport No 408

Table des matières

	Page
Résumé	iv
Abstract	v
Avertissement / Warning	vi
1. Considérations générales	1
1.1 Caractéristiques des micro-ordinateurs à utiliser	1
1.2 Logiciels utilisés dans le développement	1
1.3 Organigramme général	2
1.4 Fichiers nécessaires	3
2. Utilitaires	4
2.1 Configuration du système	4
2.2 Saisie des données	4
2.3 Traitement de fichiers	13
3. Informations générales	17
4. Fichiers "Batch"	18
5. Préparation de la série (P1.EXE)	19
5.1 Organigramme du programme	22
5.2 Contenu détaillé du programme	23
6. Études graphiques (P2.EXE)	24
6.1 Organigramme du programme	27
6.2 Contenu détaillé du programme	28
6.3 Figures du programme	29

	<u>Page</u>
7. Caractérisation de la série des mesures (P3.EXE)	35
7.1 Organigramme du programme	41
7.2 Contenu détaillé du programme	42
7.3 Figures et tableaux du programme	43
8. Série de travail, création et structure (P4.EXE)	52
8.1 Organigramme du programme	54
8.2 Contenu détaillé du programme	55
8.3 Tableaux du programme	56
9. Préparation aux tests (P5.EXE)	58
9.1 Organigramme du programme	62
9.2 Contenu détaillé du programme	63
9.3 Figures du programme	64
10. Tests	67
11. Synthèse (P6.EXE)	68
11.1 Organigramme du programme	70
11.2 Contenu détaillé du programme	71
11.3 Sortie des résultats du programme	72
12. Étude des dépassements (P7.EXE)	75
12.1 Définition des deux sous-populations	75
12.2 Choix d'une norme ou d'un seuil	75
12.3 Traitement	77
13. Structure générale des fichiers	85
13.1 Organigramme général de l'utilisation des fichiers	85
13.2 Structure des fichiers	86
13.3 Autres fichiers	92

14. Fiches techniques des tests	99
14.1 Notations	99
14.2 Application générale des tests	101
14.3 Glossaire des termes techniques	102
14.4 Structure des fiches de test	103
14.5 Fiches des tests	103
14.6 Nombre effectif d'échantillons	121
14.7 Tables complémentaires des valeurs critiques	122

RÉSUMÉ

Ce logiciel se veut une aide à la détection des tendances de la qualité de l'eau par des méthodes non-paramétriques. Il a été développé dans le cadre de l'analyse temporelle des données de la banque NAQUADAT concernant la qualité de l'eau du fleuve Saint-Laurent et de celles des effluents provenant des industries réglementées se déversant dans ce cours d'eau. Toutefois, il peut accepter comme données d'entrée toute série temporelle convenablement structurée. Son utilisation est rendu facile par sa structure graphique et interactive.

Les principales étapes du logiciel sont les suivantes:

- Partie 1: Lecture des concentrations, charges et débits mesurés.
- Partie 2: Représentation graphique des données originales.
- Partie 3: Étude des fréquences d'échantillonnage, élimination des périodes sans donnée (mois et/ou années), recherche des saisonnalités et d'une relation concentration-débit.
- Partie 4: Choix d'un intervalle de travail, d'une méthode pour remplacer les données absentes, recherche d'une persistance éventuelle.
- Partie 5: Présentation des graphiques d'inertie (double-masse et fonction CUSUM) suggérant un type de tendances (monotone ou par saut), ainsi que la date du changement éventuel. Proposition du test non-paramétrique le plus approprié, compte-tenu de la structure reconnue de la série.
- Test: Exécution du test et diagnostic.
- Partie 6: Synthèse des caractéristiques des données et des options choisies, puis interprétation paramétriques des résultats:
 - pente de la tendance
 - date de changement
 - niveaux initial et final
- Partie 7: Étude des dépassements.

ABSTRACT

This software has been designed as an aid for the detection of changes in water quality data using non-parametric methods. It has been specially developed to be used for the temporal analysis of water quality data originating from both the Saint-Laurence River concentrations included in the NAQUADAT data bank and from the effluent loadings discharged by the regulated industries into this same water body. Nevertheless this software accepts any input time-series with a proper layout. Extensive use of graphics displays and of menus makes it particularly user-friendly.

The main steps of the software are the following:

- Part 1: Reading of the measured concentrations, mass and water discharges.
- Part 2: Graphical display of the original data.
- Part 3: Study of the sampling frequencies, localisation of unsampled periods (months, years) search for seasonalities and for concentration-discharge relationships.
- Part 4: Choice of a working period and of a method for filling missing values; search for persistence.
- Part 5: Graphical display of inertial curves (mass curves and CUSUM functions) enhancing continuous or step trends as well as time of change. Suggestion of the most appropriate non-parametric trend test according to the recognized structure of the series.
- Test: Execution and diagnostic of the test.
- Part 6: Synthesis of the data characteristics as well as of the chosen options. Parametrical interpretations of the results:
 - slope of a monotonous trend
 - time of changes
 - initial and final levels of the parameters
- Part 7: Study of exceedances.

AVERTISSEMENT

Comme pour tout logiciel statistique de ce type, l'utilisateur est, par ses choix seul responsable de la validité des résultats qu'il obtient; des mises en gardes répétées sont données lors du déroulement des étapes critiques. Face à des plans d'échantillonnage historiques changeant, le choix d'un intervalle de travail approprié nécessite, d'une part, de s'assurer que les données originales recueillies sont représentatives du phénomène considéré et, d'autre part que l'intervalle choisi ne crée pas une quantité inconsidérée d'information fictive.

WARNING

As for any statistical software of this kind, the user by his choices is solely responsible of the validity of the obtained results. Numerous notices of caution are provided at the critical steps of the procedure. With the existence of varying historical sampling schemes, the choice of an appropriate working interval should be made firstly by considering the representativity of the original data with regards to the sampled phenomenon, and secondly, by insuring that this choice is not introducing some inconsiderate amount of fictitious information.

1. Considérations générales

1.1 Caractéristiques des micro-ordinateurs à utiliser

Le micro-ordinateur d'exploitation doit être un ordinateur personnel compatible de type PC, XT ou AT fonctionnant sous le système d'exploitation DOS 2.0 ou plus; il doit comporter:

- 512 kb de mémoire vive (au moins);
- une plaque d'interface graphique CGA, EGA ou ATT 6300B-OLIVETTI;
- un écran monochrome ou couleur à capacité graphique;
- 2 lecteurs de disquettes ou un disque dur et un lecteur de disquettes; (360 kb 5½ po, double face-double densité)
- une imprimante à aiguilles à capacité graphique;
- un modem compatible HAYES permettant la communication avec un ordinateur éloigné à 300,1200 ou 2400 bauds;
- un coprocesseur 8087 ou 80287 n'est pas nécessaire, mais accélère les calculs s'il est présent.

1.2 Logiciels utilisés dans le développement

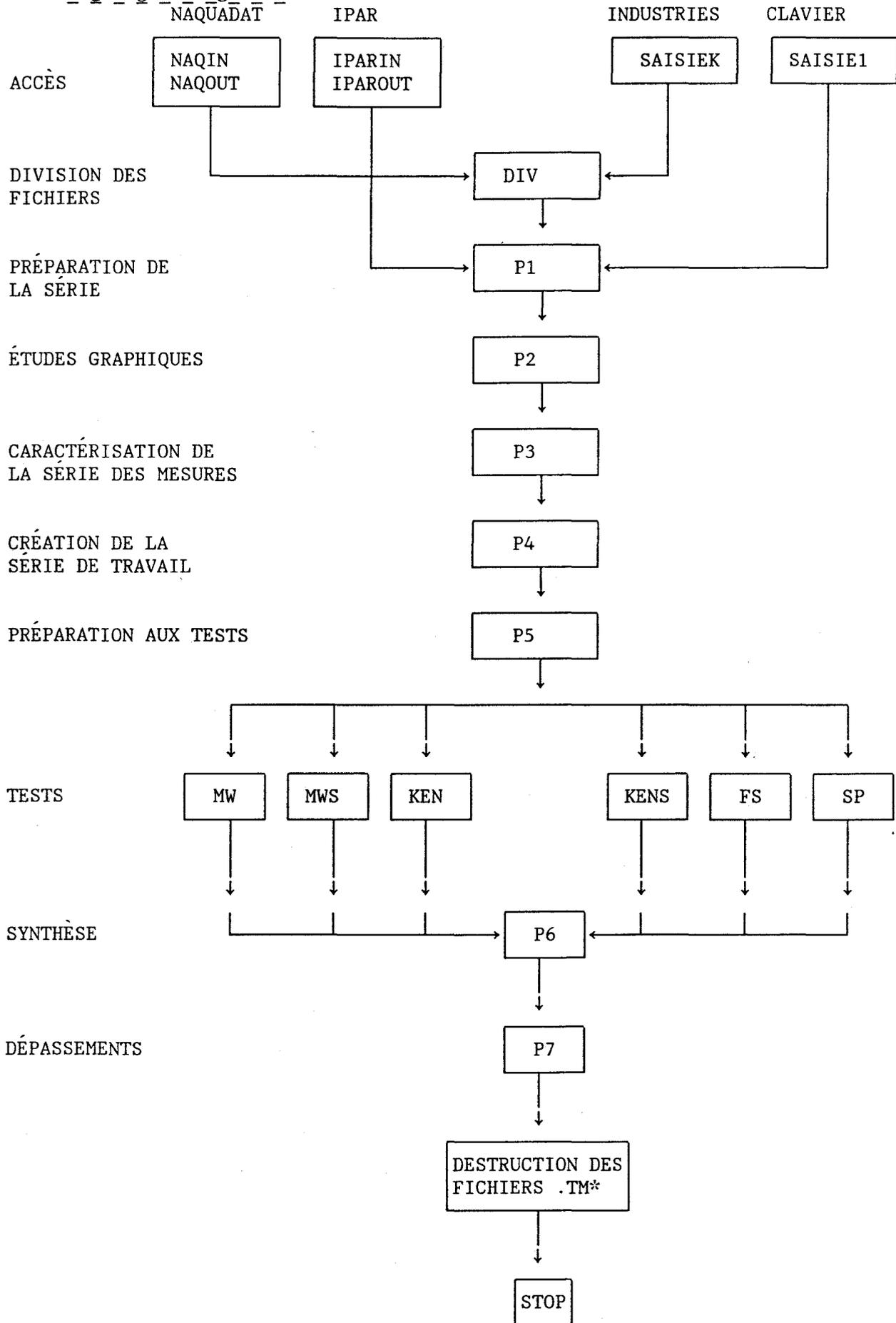
La plupart des programmes ont été développés en FORTRAN 77 et compilés avec le logiciel MICROSOFTTM version 3.31.

La partie graphique interactive a été effectuée en FORTRAN 77 avec le logiciel Grafmatic de MICROCOMPATIBLESTM.

Les fichiers de communication ont été écrits de façon à être compatibles avec le logiciel Crosstalk 16, version 3.5 de MICROSTUFTM.

En ce qui concerne le développement des programmes eux mêmes et la manipulation des fichiers, deux éditeurs faisant appel aux mêmes commandes ont été utilisés. Pour les fichiers de longueur inférieure à 50 kb, on a utilisé Sidekick (1:52A) de BORLAND INTERNATIONALTM et pour les fichiers plus longs Wordstar (3.30) de MICROPRO INTERNATIONALTM.

1.3 Organigramme général



1.4 Fichiers nécessaires

Les fichiers suivants doivent être présents lors de l'utilisation du logiciel:

a) fichiers exécutables, (.EXE)

- Parties du logiciel : P1.EXE, P2.EXE, P3.EXE, P4.EXE, P5.EXE, P6.EXE, P7.EXE
- Tests de logiciel: MW.EXE, MWS.EXE, KEN.EXE, KENS.EXE, SP.EXE, KENS.EXE
- Autres utilitaires: CONF.EXE, DIVISE.EXE, SAISIE.EXE.

b) fichiers de données

- tables de valeurs critiques pour certains tests:
FISCHER.TAB, CHIDEUX.TAB, TESTRHO.TAB, STUDENT.TAB, FS2.TAB
- fichiers contenant les séries de concentrations, de charges ou de débits. Ces fichiers sont obtenus à l'aide des utilitaires.
- 1 fichier contenant les codes NAQUADAT et les extensions correspondantes pour DIVISE.EXE(CODES.NAQ).

N.B. Les fichiers de valeurs critiques (.TAB) ainsi que le fichier CODES.NAQ sont présentés dans la partie 13.3.

2. Utilitaires

Afin de permettre une utilisation plus efficace du logiciel, quelques utilitaires sont fournis avec le logiciel. Voici une brève description de ces programmes complémentaires.

2.1 Configuration du système

CONF.EXE

Comme le logiciel doit pouvoir être adapté à des microordinateurs de type PC équipés différemment, il existe un programme de configuration CONF.EXE qui permet au logiciel de s'adapter au système de l'utilisateur. Ce programme pose des questions concernant la définition des écrans, les modes couleurs utilisées, les symboles pour les graphiques et la définition des interfaces utilisés. Les programmes sont adaptés aux modes CGA (640 x 200), EGA (640 x 350) et ATT (640 x 400); ils fonctionnent aussi en mode 320 x 200, mais certains titres sont alors superposés, car le mode texte correspondant à 40 x 25 ne donne que peu de latitude pour commenter les figures.

Le programme CONF.EXE crée un fichier SETUP.PC qui doit être présent dans l'unité de défaut lors de l'exécution du logiciel.

2.2 Saisie de données

Les données à saisir peuvent provenir de trois sources:

- des données de qualité de l'eau du Saint-Laurent de la Banque NAQUADAT et des données de débit du Fleuve de la banque IPAR.
- des données de débits et de charges rejetés par les industries réglementées.
- des données temporelles de quantité et de qualité d'origine quelconque.

A) Données concernant l'eau du Saint-Laurent

Les données de qualité sont localisées sur la banque NAQUADAT dans l'ordinateur CDC situé au Centre de calcul du Ministère de l'Énergie des Mines et des Ressources à Ottawa; les données de niveau et de débit du Fleuve se trouvent au même endroit sur une banque nommée HYDAT. Nous avons écrit un certain nombre de programmes compatibles au logiciel de communication CROSSTALK permettant la configuration du modem, le dépôt des requêtes et la saisie des fichiers de données de façon quasi-automatique. Ces fichiers devront être édités selon le matériel de l'utilisateur et les requêtes effectuées, une connaissance du logiciel CROSSTALK n'est pas nécessaire cependant, l'utilisateur doit avoir une certaine connaissance du logiciel IPAR et des banques de données NAQUADAT et HYDAT afin de pouvoir effectuer les changements qui lui permettront d'avoir les données qu'il désire.

Six programmes sont présentés: le premier, EMR.XTK présente la configuration nécessaire pour communiquer avec l'ordinateur serveur. Les programmes NAQIN.XTK, NAQOUT.XTK, IPARIN,XTK et IPAROUT,XTK sont identiques à ce programme et ne sont pas présentés. Leur présence est cependant obligatoire car le programme NAQIN.XTS est effectué après NAQIN.XTK, NAQOUT.XTS après NAQOUT.XTK etc. Pour plus de renseignements l'utilisateur peut se référer au manuel d'utilisation de CROSSTALK.

Les cinq autres programmes effectuent les opérations nécessaires à une acquisition automatique des données. Ces programmes devront être édités afin d'obtenir les données désirées.

Programme 1: EMR.XTK

Ce programme configure le modem en fonction de la sortie série utilisée, de la vitesse de communication et du protocole de l'ordinateur-serveur. Les programmes NAQIN.XTK, NAQOUT.XTK, IPARIN.XTK et IPAROUT.XTK sont identiques à EMR.XTK.

```
Name      NA1200Q
Number    6472691
ACcept    Everything
ANswback  Off
APrefix   ATSO=1{
ATten     Esc
BReak     End
SWitch    Home
CWait     None
LWait     None
SC        N Cb
DEbug     Off
DPrefix   ATVOX4{ }ATDT
DSuffix   {
EMulate   None
EPath     ""
Filter    -----+++++-----
INfilter  On
LFauto    Off
MObde     Call
POrt      2
PWord     ""
RDials    10
TImmer    On
TUrnarnd  Enter
BKsize    1
DNAMES    200
CApture   Off
COmmand   ETX (eC)
FLOw      eS/eQ
PARity    EVEN
DAta      7
DUplex    Full
OUtfilter On
PMode     2 (DOS)
PRinter   Off
SPEed     1200
STop      1
TABex     Off
BLankex   Off
SC        H Yr
UConly    Off
FK 1 @QUIT{
FK 2 ""
FK 3 ""
FK 4 ""
FK 5 @CApture /{
FK 6 @PRinter /{
FK 7 @TYpe{
FK 8 @CApture <24{
FK 9 @SNap{
FK 10 @SNap 24{
FK S1 ""
```

```
FK S2 ""
FK S3 ""
FK S4 ""
FK S5 ""
FK S6 ""
FK S7 ""
FK S8 ""
FK S9 ""
FK S10 ""
FK C1 ""
FK C2 ""
FK C3 ""
FK C4 ""
FK C5 ""
FK C6 ""
FK C7 ""
FK C8 ""
FK C9 ""
FK C10 ""
FK A1 ""
FK A2 ""
FK A3 ""
FK A4 ""
FK A5 ""
FK A6 ""
FK A7 ""
FK A8 ""
FK A9 ""
FK A10 ""
SC L Wr
GO q45/45
```

Programme 2: EMR.XTS

Ce programme permet la communication automatique avec l'ordinateur de Énergie, Mines et Ressources (EMR), donne le nom de l'utilisateur et son mot de passe et ouvre le dialogue.

```
WAIT QUIET 40
REPLY .{ DO
WAIT QUIET 30
REPLY 85701143,/MA/{ DO [1] [2]
WHEN "PLS LOGIN SVP" REPLY LOGIN,ICKXX,X87QUER { :DO
WAIT MANUAL
DUPLEX HALF
```

[1]: Nom de l'utilisateur*
[2]: Mot de passe*

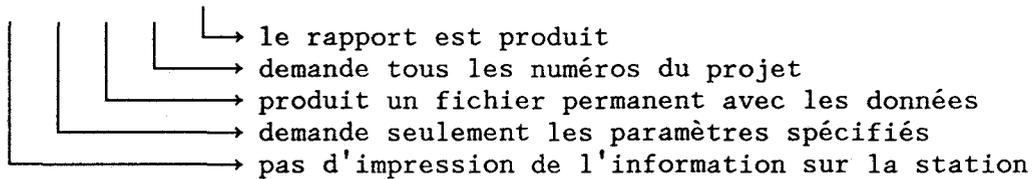
*: peut être changé par les nouveaux usagers à l'aide de Sidekick ou autre éditeur de texte.

Programme 3: NAQIN.XTS

Ce programme permet de soumettre une demande de données de qualité à la banque NAQUADAT selon son protocole de recherche interactive.

```
WAIT QUIET 40
REPLY .{ DO
WAIT QUIET 30
REPLY 85701143,/MA/{ DO           [1]   [2]
WHEN "PLS LOGIN SVP" REPLY LOGIN,ICKXX,X87QUER { :DO
WAIT MANUAL
DUPLEX HALF
WAIT STRING "COMMAND-"
REPLY TXTEDIT{
WAIT STRING "++"
REPLY AUTO FROM 10{
WAIT CHAR "="
[3] REPLY IDACL,CM210000,T1000,IO10000,STWDJ.{
WAIT CHAR "="
[4] REPLY ACCOUNT,55510.{
WAIT CHAR "="
REPLY PAUSE. CLOCK TIME APPROX 1 HOUR {
WAIT CHAR "="
[5] REPLY BEGIN,NAQDR,NAQPROC/WQB,SETNM,OUTF=ESS,IDOUT=CLA.{
WAIT CHAR "="
REPLY *EOR {
WAIT CHAR "="
[6] REPLY READ      NAQCENT {
WAIT CHAR "="
[7] REPLY 0 1 1 0 0{
WAIT CHAR "="
[8] REPLY 00QU020B9016.{
WAIT CHAR "="
[9] REPLY 750101 860101 {
WAIT CHAR "="
[10] REPLY 02041L ,02041M {
WAIT CHAR "="
REPLY ={
WAIT STRING "++"
REPLY SAVE,NAQ,NS {
WAIT STRING "++"
REPLY BYE {
WAIT STRING "COMMAND-"
REPLY BATCH,NAQ,INPUT,HERE {
WAIT STRING "COMMAND-"
REPLY LOGOUT {
QUIT
```

- [1]: Nom de l'utilisateur*
- [2]: Mot de passe*
- [3]: IDACL* : Nom de l'utilisateur
CM210000*: "CORE MEMORY" de 210K
T1000* : "CPU TIME" en secondes octales de 1000 sec.
IO10000* : "IO TIME" en secondes octales de 10000 sec.
STWDJ* : Priorité
- [4]: 55510* : Numéro de compte NAQUADAT
- [5]: OUTF=ESS*: Identification du fichier permanent contenant les
IDOUT=CLA* résultats (facultatif).
- [6]: NAQCEN*: Nom de la banque de données (NAQCEN contient Québec et
Ontario)
- [7]: 0* 1* 1* 0* 0*



- [8]:00QU020B9016.*: Numéro de la station
- [9]:750101* 860101*: Date (AAMMJJ) de départ et de fin de la série que l'on
veut étudier
- [10]:02041L*, 02041M*: Numéro du paramètre et du paramètre de "backup", séparés
par une virgule (on peut avoir plus d'un code de
"backup")

* Toutes ces valeurs peuvent être changées à l'aide d'un éditeur de texte afin d'effectuer des demandes différentes.

Pour plus d'information sur les numéros [5] à [10], l'utilisateur peut se référer au guide d'utilisation de la banque NAQUADAT.

Programme 4: NAQOUT.XTS

Ce programme permet ultérieurement la recherche du nom du fichier résultat et son transfert vers le micro-ordinateur par l'intermédiaire de la commande

CAPTURE ON - le format de saisie est celui des rapports détaillés.

WAIT QUIET 40

REPLY .{ DO

WAIT QUIET 30

REPLY 85701143,/MA/{ DO [1] [2]

WHEN "PLS LOGIN SVP" REPLY LOGIN,ICKXX,X87QUER { :DO

WAIT MANUAL

DUPLEX HALF

WAIT STRING "COMMAND-"

[3] ASK @F1 NOM DU REMOTE OUTPUT FILE :

REPLY BATCH,@F1,LOCAL {

WAIT STRING "COMMAND-"

REPLY CONNECT,OUTPUT {

CAPTURE ON

WAIT STRING "COMMAND-"

REPLY COPYSBF,@F1,OUTPUT {

WAIT STRING "COMMAND-"

REPLY LOGOUT {

[4] WRITE TOTO2.DAT

CAPTURE OFF

QUIT

[1]: Nom de l'utilisateur*

[2]: Mot de passe*

[3]: A ce moment, CROSSTALK demande à l'utilisateur le nom du fichier "REMOTE" qui doit apparaître à l'écran. Ce fichier aura un nom de la forme ICKXX??, soit le nom de l'utilisateur plus deux caractères non définis à l'avance. Il est donc impossible d'éditer le programme avant l'exécution pour changer le nom du fichier "REMOTE".

[4]: TOTO2.DAT*: nom du fichier qui sera créé sur le micro-ordinateur

* Ces valeurs peuvent être changées à l'aide d'un éditeur de texte par l'utilisateur.

Programme 5: IPARIN.XTS

Ce programme permet de requérir la mise en fichier de données de débits du fleuve Saint-Laurent.

```
WAIT QUIET 40
REPLY . { DO
WAIT QUIET 30
REPLY 85701143,/MA/{ DO          [1]    [2]
WHEN "PLS LOGIN SVP" REPLY LOGIN,ICKXX,X87QUER { :DO
WAIT MANUAL
DUPLEX HALF
[3] WHEN "COMMAND-" REPLY BEGIN,,IPAR/WRB,SUP { :DO
WAIT MANUAL
WAIT STRING "NAME :"
[4] REPLY CLUIS,D {
WAIT CHAR ":"
[5] REPLY UNIVERSITE DU QUEBEC {
WAIT CHAR "?"
REPLY N {
WAIT CHAR ":"
REPLY 01 {
WAIT STRING "Y/N: ?"
REPLY N {
WAIT STRING "Y/N: ?"
REPLY N {
WAIT QUIET 50
[6] REPLY 020A016-020A016 1950-1986 {
REPLY END {
WAIT STRING "Y/N: ?"
REPLY N {
WAIT STRING "DEVICE(S) :"
REPLY D {
WAIT STRING "CHARACTERS :"
[7] REPLY DEB0165086 {
WAIT STRING "CHARACTERS :"
[8] REPLY CLA {
WAIT STRING "INFINITE) :"
REPLY 999 {
WAIT STRING "(H) :"
REPLY H {
WAIT STRING "Y/N: ?"
REPLY N {
WAIT CHAR "?"
REPLY N {
WAIT STRING "Y/N: ?"
REPLY Y {

WAIT STRING "Y/N: ?"
REPLY N {
WAIT STRING "NEW LINE)"
REPLY END {
WAIT STRING "COMMAND-"
REPLY LOGOUT {
QUIT
```

- [1]: Nom de l'utilisateur*
- [2]: Mot de passe*
- [3]: Commande d'appel du programme interactif IPAR
- [4]: CLUIS*, D*: nom du requérant
- [5]: Université du Québec* : organisme du requérant
- [6]: 020A016* - 020A016* : numéro de station de départ et de fin
1950*-1986* : année de départ et de fin
- [7]: DEB0165086* : nom du fichier permanent qui sera créé
- [8]: CLA* : caractères pour "ID" du fichier permanent

* ces valeurs peuvent être changées par l'utilisateur à l'aide d'un éditeur de texte

N.B. Pour plus d'information sur la signification des réponses, on doit se référer au manuel d'utilisation du logiciel IPAR.

Programme 6: IPAROUT.XTS

Ce programme permet ultérieurement le transfert du fichier résultat grâce à la commande CAPTURE ON.

```
WAIT QUIET 40
REPLY .{ DO
WAIT QUIET 30
REPLY 85701143,/MA/{ DO           [1] [2]
WHEN "PLS LOGIN SVP" REPLY LOGIN,ICKXX,X87QUER { :DO
WAIT MANUAL
DUPLEX HALF
WAIT STRING "COMMAND-"
[3] REPLY ATTACH, TOTO, DEB0165086, ID=CLA {
WAIT STRING "COMMAND-"
REPLY CONNECT,OUTPUT {
WAIT STRING "COMMAND-"
[4] CAPTURE TOTO. DAT /E
REPLY COPYSBF,TOTO,OUTPUT {
WAIT STRING "COMMAND-"
CAPTURE OFF
REPLY LOGOUT {
WAIT QUIET 20
QUIT
```

- [1]: Nom de l'utilisateur*
- [2]: Mot de passe*
- [3]: DEB0165086*,ID=CLA* : doit correspondre au nom du fichier permanent
identifié dans IPARIN.XTS
- [4]: TOTO.DAT* : nom du fichier qui sera créé sur le micro-ordinateur

*: Ces valeurs peuvent être changées par l'utilisateur à l'aide d'un éditeur de texte

N.B. Ce programme peut servir pour transférer tout fichier permanent présent sur l'ordinateur de E.M.R., on n'a alors qu'à éditer les lignes [3] et [4] de façon pertinente.

B) Données concernant les rejets d'industries (SAISIE.EXE)

Les données de rejet des industries n'étant pas, à l'origine, sur support informatique, nous avons écrit un programme permettant une saisie simple de ces données et une écriture pouvant être lu facilement par le programme de division. Ce programme permet également la saisie d'une seule série sous un format qui peut être lu directement par pl.

L'utilisateur est guidé à l'aide de menus et de questions lui permettant de:

- Définir la présence ou non des dates (la présence de dates est obligatoire pour un traitement par le logiciel).
- Éliminer un enregistrement mal saisi.
- Annexer des données à un fichier déjà existant.

2.3 Traitement de fichiers

DIVISE.EXE

Les fichiers NAQUADAT et les fichiers industriels contiennent des données de concentrations ou de charges pertinentes à plusieurs paramètres. Ce programme effectue la division automatique de ces fichiers en fichiers ne contenant qu'un seul paramètre.

Pour exécuter le programme l'utilisateur n'a qu'à taper: DIVISE lorsque dans l'unité de disque contenant le programme DIVISE.EXE. Au début de l'exécution, on demande à l'utilisateur le nom du fichier à diviser, ex.: DONNEES.DAT. Trois types de fichiers peuvent être traités. L'utilisateur devra choisir entre:

1. Fichier type NAQUADAT normal
2. Fichier type NAQUADAT avec des <CR> après 80 caractères
3. Fichier type "industries"

Voyons les caractéristiques de chacun de ces types de fichier et les traitements qui sont effectués dans chaque cas.

1- Fichier du type NAQUADAT normal

Les noms des fichiers de sortie seront construits automatiquement à partir des renseignements compris dans le fichier d'entrée. Ce fichier devra donc contenir en première ligne; NP: le nombre de paramètres (NP doit être ≤ 25) ainsi que les codes analytiques de chacun des NP paramètres. Le format de lecture de cette première ligne est:

(3X,I2,1X,A5,3X,A5,3X,...,A5)

Comme seulement les cinq premiers caractères des codes analytiques sont lus, le programme ne fera donc pas de distinction entre les codes 02041L, 02041L2 et 02041S. Les codes sont utilisés pour créer l'extension des fichiers de sortie. Les codes reconnus et les extensions correspondantes sont contenus dans le fichier CODES.NAQ. Ce fichier peut facilement être édité afin d'enlever ou d'ajouter des codes ainsi que les extensions correspondantes. On peut également ajouter, en bout de lignes, des commentaires permettant d'identifier exactement le paramètre dont il est question.

Normalement, il apparaît une ligne blanche après la première ligne contenant les codes analytiques. Cependant, les fichiers peuvent avoir été édités de sorte que cette ligne blanche n'est pas présente. Dans ce cas, lorsque le programme demandera d'identifier le fichier d'entrée:

- 1) Fichier original en provenance de NAQUADAT
- 2) Fichier manipulé à l'aide d'un éditeur de texte

il faudra choisir la seconde option.

Les lignes suivantes du fichier d'entrée doivent avoir le format NAQUADAT classique et seront lues de la façon suivante:

- Lecture d'une variable de 12 caractères identifiant la station. Les six derniers caractères sont utilisés pour construire les noms des fichiers de sortie.
- Lecture d'une variable de 24 caractères inutiles à la division.
- Lecture des valeurs des NP paramètres sous un format libre (les valeurs n'ont donc qu'à être séparées par un espace).

Il est bon de noter que le fichier d'entrée peut contenir des mesures de plusieurs stations différentes. Le programme changera les fichiers de sortie lorsqu'il rencontrera une nouvelle station.

EXEMPLE:

Supposons le fichier d'entrée suivant:

```
2 02041S 10301L
00QU02MC9202 ..... 90.0 7.2
00QU02MC9202 ..... 91.0 7.1
.
.
.
00QU02MC9202 ..... 93.0 7.3
00QU02MC9205 ..... 98.1 6.4
00QU02MC9205 ..... 99.4 6.5
.
.
.
00QU02MC9205 ..... 100.1 6.2
```

En utilisant le programme *divise*, on créera automatiquement les 4 fichiers MC9202.CON, MC9202.PH, MC9205.CON et MC9205.PH qui contiendront respectivement les mesures de conductivité et de pH pour les stations 00QU02MC9202 et 00QU02MC9205.

2- Fichier de type NAQUADAT avec des <CR> après 80 caractères

Cette option effectue les mêmes opérations que l'option (1). Elle est offerte car dans certains cas, le transfert de fichiers d'un type d'ordinateur à un autre peut amener la présence de <CR> (retours de chariot) après 80 caractères.

3- Fichier de type "industries"

Cette option permet le traitement de presque tous les types de fichiers qui ne viennent pas de NAQUADAT. On demande d'abord à l'utilisateur le nombre de colonnes comprises dans le fichier d'entrée (date comprise). L'utilisateur est ensuite guidé à l'aide de menus et d'explications lui permettant de:

- Localiser l'endroit où est la date dans le fichier. À noter que l'absence de date ne permettra pas une utilisation dans le logiciel même si on peut réaliser DIVISE. Cette option est introduite pour une utilisation externe au logiciel lorsque les données sont équidistantes.
- Localiser la colonne des débits s'ils sont présents.
- Transformer les concentrations en mg/l les charges en Kg/j et les débits en m³/sec afin d'avoir concordance dans les résultats du logiciel.
- Permettre l'écriture ou non des dates dans les fichiers de sortie. (Les dates sont cependant obligatoires pour un traitement par le logiciel).

Les noms des fichiers de sortie sont finalement demandés à l'utilisateur. Celui-ci n'est pas dans l'obligation de créer des fichiers pour tous les paramètres compris dans le fichier de départ. Il peut par exemple ne choisir qu'un seul paramètre et ne créer qu'un seul fichier de sortie.

3. Informations générales

On appelle logiciel, l'ensemble des programmes visant à détecter des tendances. Il s'agit des programmes P1, P2, P3, P4, P5, P6, MW, MWS, KEN, KENS, FS, SP et P7. Les autres programmes fournis sont appelés utilitaires: CONF, DIVISE, SAISIE ainsi que l'ensemble des programmes associés au logiciel CROSSTALK.

Le logiciel est, dans sa forme originale, dimensionné à 1300. Pour des séries contenant plus de 1300 valeurs deux choix s'offre à l'utilisateur:

- Réduire, à l'extérieur du logiciel, la longueur de la série.
- Augmenter à l'intérieur du logiciel (programmes sources ".FOR") le dimensionnement des variables (paramètre MAXI). Des essais ont cependant montré que l'utilisation d'un dimensionnement plus grand que 1300 peut amener un problème d'espace mémoire.

Le logiciel a été développé avec l'idée qu'il sera utilisé sur un disque dur et bien qu'il puisse être utilisé sur un micro-ordinateur contenant deux lecteurs de disquettes, une utilisation efficace demande la présence d'un disque dur. Afin d'utiliser le logiciel sur un micro contenant deux lecteurs de disquettes, on doit:

- 1- mettre la disquette de données dans le lecteur par défaut;
- 2- appeler les autres programmes du lecteur par défaut;
- 3- tous les fichiers intermédiaires seront créés sur le lecteur par défaut.

Exemple:

- 1- mettre la disquette de données dans le lecteur A;
- 2- choisir A comme lecteur par défaut si ce n'est pas déjà le cas:

A:←|

- 3- à la réponse A> on répond:

B:CONF

si l'on doit créer le fichier SETUP.PC ou encore:

B: P1

si l'on veut débiter l'utilisation du logiciel

On appellera de cette façon, toutes les parties du logiciel en s'assurant que la disquette correspondante se trouve dans lecteur B.

4. Fichiers "Batch"

Afin de permettre une exécution rapide et automatique des différentes parties du programme une série de fichiers "batch" (.BAT) ont été créés. Il s'agit de:

DETECT.BAT

PP1.BAT

PP2.BAT

PP3.BAT

PP4.BAT

PP5.BAT

chacun de ces programmes permet de repartir à un endroit différent du logiciel, il faut cependant que l'exécution précédente ait été déjà réalisée jusqu'à un niveau plus avancé que l'endroit du nouveau départ. Exemple: on peut reprendre à PP3 après une exécution complète du logiciel mais on ne peut pas reprendre à PP3 après une exécution arrêtée à P1.

Comme les programmes "batch" appellent les fichiers exécutables (P1, P2, ..., P6 ainsi que les tests) il est donc possible de ne pas utiliser ces fichiers "batch" et de ne travailler qu'avec les fichiers ".EXE"; cependant, il est préférable pour l'utilisateur peu habitué d'utiliser les fichiers ".BAT", car les fichiers intermédiaires pouvant créer des conflits sont alors automatiquement éliminés.

5. Préparation de la série (Pl.EXE)

Ce programme effectue les opérations suivantes:

- . lecture complète du fichier d'entrée (concentrations ou charges)
- . localisation des débits si ils existent
- . élimination des dates où le paramètre étudié n'a pas été analysé
- . remplacement de la date du prélèvement par le nombre de jours depuis le premier prélèvement de la série
- . création du second paramètre de qualité (concentration ou charge) à partir de la série des débits si elle existe
- . écriture d'un fichier de travail .TMP pour utilisation ultérieure

Généralités

FICHER D'ENTRÉE

Le fichier d'entrée peut contenir des charges ou des concentrations, les fichiers de données originaux fournis avec le logiciel contiennent toujours les concentrations. Le fichier d'entrée doit avoir le format FORTRAN suivant: (12X, I2, 1X, I2, 1X, I2, 16X, F12.6, F12.6). On doit donc avoir:

- 12 espaces
- un entier de longueur 2 contenant l'année
- un espace
- un entier de longueur 2 contenant le mois
- un espace
- un entier de longueur 2 contenant le jour
- 16 espaces
- un réel de longueur 12 avec 6 chiffres après le point contenant la charge ou la concentration
- un réel de longueur 12 avec 6 chiffres après le point contenant le débit. Ce dernier étant facultatif on doit laisser cet espace vide si les débits ne sont pas tirés de ce fichier.

Voici quelques exemples de fichiers qui seront lus correctement par le logiciel:

79-04-19	1.900000
79-05-10	2.300000
79-06-07	1.900000
79-06-29	2.000000
OOQU02MC920279-04-191440EST 805 25	7.900000
OOQU02MC920279-05-191550EST 805 25	8.400000
OOQU02MC920279-06-071200EST 805 25	8.200000
OOQU02MC920279-06-293400EST 805 25	8.200000
OOQU02MC920279-08-281300EST 805 25	8.100000

79-04-19	1.900000	.212341
79-05-10	2.300000	.198743
79-06-07	1.900000	.207432
79-06-29	2.000000	.198672

On voit donc que les espaces peuvent être laissés blancs ou non puisque ces champs ne sont pas lus. Ce format a été choisi à cause de sa conformité avec les sorties obtenues de la banque NAQUADAT.

UNITÉS

Pour avoir correspondance avec les unités présentées dans le logiciel, les concentrations doivent être en mg/l, les charges en kg/J et les débits en m³/sec.

FICHER DE DÉBIT

Si on utilise un fichier de débits journaliers, il doit avoir le format de sortie du programme interactif IPAR:

Q020A016955	11	8640	8830	8980	8610	8580	8380	8720	8100	8440	8720	31
Q020A016955	12	8520	8520	8610	8610	8470	8920	9090	8670	8780	8830	
Q020A016955	13	8810	8640	8830	9320	9090	8950	8920	8550	8520	8610	8830
Q020A016955	21	8500	8440	8040	8040	8100	8350	8810	8300	8210	7960	28
Q020A016955	22	8410	8410	7960	7930	8130	8210	8040	8160	8100	7990	
Q020A016955	23	8100	8100	8070	8160	8270	8100	8180	8240	-1111	-1111	-1111

où, pour la première ligne:

Q020A016 = numéro de station

955 = identifie l'année

11 = identifie le mois et la décade

ex.: 11 = 1er mois, 1er dix jours

8640 = valeur du débit journalier 1er jour de la décade

8830 = valeur du débit journalier 2e jour de la décade

.

.

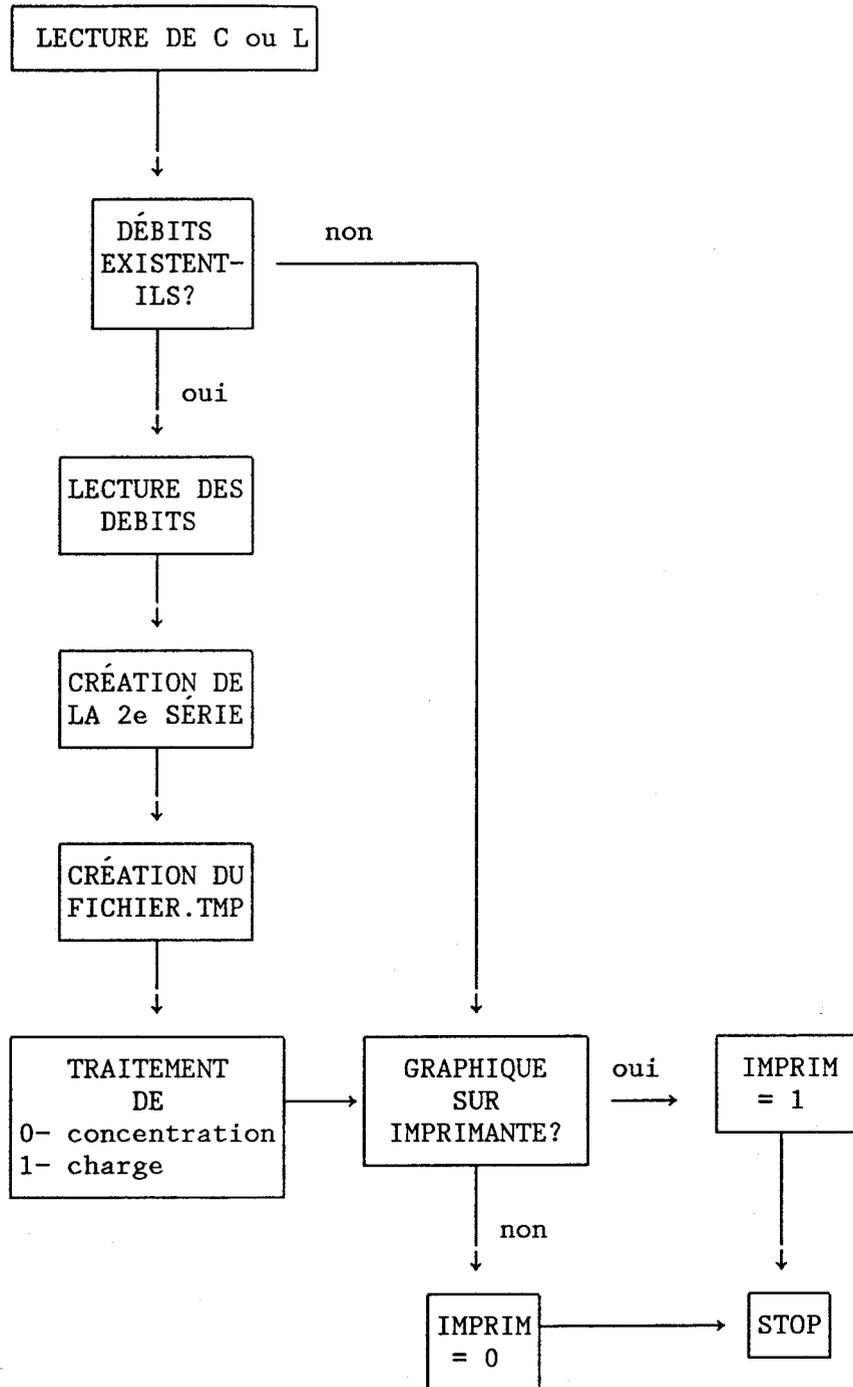
.

8720 = valeur du débit journalier 10e jour de la décade

31 = nombre de jours dans le mois

Le fichier DEB.ESS fourni avec le logiciel, contient les débits journaliers de 1955 à 1985 pour la station Q020A016 et son format est celui produit par le logiciel IPAR.

5.1 Organigramme du programme P1.EXE



5.2 Contenu détaillé du programme P1.EXE

ACTION	RESTRICTION	SOUS-ROUTINES UTILISEES
Lecture de C ou L	Les dates relatives aux concentrations ou aux charges doivent être en ordre chronologique. Les données négatives sont considérées comme données manquantes et ne sont pas retenues.	
Lecture des débits	Si les débits ne sont pas dans le même fichier que C ou L et qu'on désire les utiliser, un fichier de débits journaliers standard (IPAR) doit être disponible.	TROUVDEB
Création de la 2 ^e série	Seulement si les débits sont utilisés.	
Création du fichier TMP		
Choix du traitement de C ou de L	Il n'est pas possible de traiter C et L à la fois.	
Choix d'envoyer ou non les graphiques à l'imprimante	Dans l'affirmative, l'imprimante utilisée doit supporter le mode graphique.	

6. Études graphiques (P2.EXE)

Ce programme lit le fichier de travail précédemment créé et effectue des présentations graphiques; il présente d'abord l'évolution temporelle du paramètre étudié et suggère l'élimination des maxima et minima, qui, s'ils sont aberrants peuvent biaiser les interprétations graphiques. On doit noter que pour la détection elle-même, le choix de méthodes non-paramétriques limite l'impact de ces valeurs.

Ensuite il propose de tracer:

- . les courbes double-masse
- . les fonctions CUSUM

GÉNÉRALITÉS

VALEURS ABERRANTES

L'élimination des valeurs aberrantes vise surtout à ne pas biaiser les résultats des analyses graphiques ainsi que différentes analyses paramétriques effectuées à l'intérieur du logiciel. En ce qui concerne la détection de tendance, l'utilisation de tests non-paramétriques permet d'avoir un résultat stable malgré la présence de telles valeurs.

Lorsqu'une valeur est éliminée comme aberrante par l'utilisateur, cette valeur ne sera plus utilisée par la suite dans le logiciel et la seule façon de faire revenir cette valeur consiste à répartir à P1 avec le fichier original.

COURBE DOUBLE-MASSE

Les courbes double-masses présentent, sous forme de sommes accumulées, l'évolution des paramètres. Pour une courbe double-masse paramètre vs temps, on a en ordonnée au temps t^* : $\sum_{t=0}^{t^*} x_t$, où x_t est la valeur du paramètre au temps t . On aura en abscisse le rang du temps t^* en nombre de jours depuis la première observation mesurée.

Dans le cas d'une courbe double-masse paramètre vs débit, on aura en abscisse au temps t^* : $\sum_{t=0}^{t^*} Q_t$. L'ordonnée sera calculée comme pour les doubles-masses paramètre vs temps.

Étant donné la non-équidistance des observations, les courbes doubles-masses doivent surtout être utilisées ici comme méthode exploratoire de la possibilité de présence d'une tendance, leur utilisation dans P5 avec des données équidistantes sera plus représentative de la nature du phénomène étudié.

Que faut-il chercher?

A) Double-masse paramètre vs temps

Les figures 2.2 et 2.3 présentent de telles courbes doubles-masses. La droite partant du point (0,0) et se rendant au point supérieur droit du cadre, est appelée droite de la moyenne. On doit surtout regarder de tels graphiques comme une vue d'ensemble des changements de pentes. Ainsi, lorsqu'un groupement de points successifs semble former une droite avec une pente plus grande que la pente de la droite moyenne, on peut alors conclure que cet ensemble de points a une moyenne supérieure à la moyenne générale. Une pente plus faible permettra aussi de conclure que l'ensemble de points correspondant, a une moyenne inférieure à la moyenne générale. On peut alors facilement voir l'utilité de telles courbes pour la détection de tendances sur la moyenne.

On peut également voir, sur les figures 2.2 et 2.3, des segments de part et d'autre de la droite de la moyenne. Ces "rails" sont situés à deux écart-types de la droite de la moyenne et chaque écart-type est calculé exclusivement à l'aide des points situés du côté correspondant de la droite moyenne. Ces "rails" si ils sont éloignés de la droite moyenne suggèrent des grandes excursions et permettent de détecter un écart important de la courbe double-masse par rapport à la droite de la moyenne. Un "rail" est confondu avec la droite de la moyenne si tous les points sont situés d'un seul côté de cette droite.

Si aucune tendance sur la moyenne n'est présente, les points de la courbe double-masse devraient être situés de part et d'autre de la droite de la moyenne avec un cheminement "aléatoire".

B) Double-masse paramètre vs débit

Le principe de telles courbes double-masse est le même que pour le cas paramètre vs temps. Cependant, on cherche aussi à savoir si une tendance détectée vs le temps ne pourrait pas être introduite par un effet de débit. On recherche donc un "patron" différent de celui de la courbe double-masse paramètre vs temps.

FONCTION CUSUM

Les fonctions CUSUM doivent être étudiées de la même façon que les courbes doubles-masses en ce qui concerne les changements de pentes de points successifs. On calcule la fonction CUSUM de la façon suivante:

$$\text{CUSUM}(x_t) = \sum_{j=1}^t x_j - j \cdot \bar{x}$$

et on trace le graphique de la fonction CUSUM en fonction du temps t .

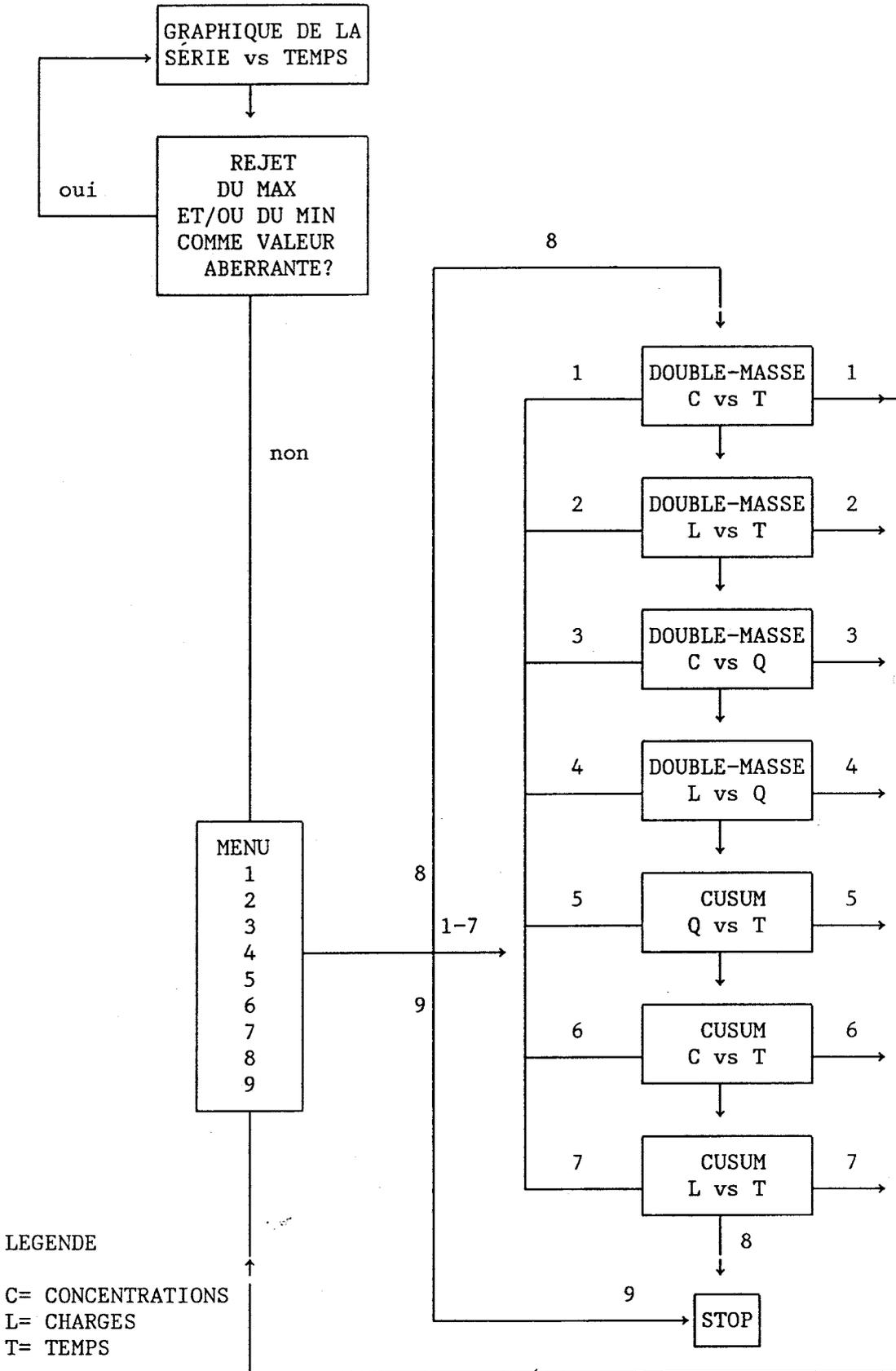
Que faut-il chercher?

La fonction CUSUM (sommés cumulatives) effectue donc, en quelque sorte, une rotation de la droite de la moyenne des figures 2.2 et 2.3 afin de l'amener à l'horizontale. Les écarts à la droite de la moyenne sont alors beaucoup plus visibles puisqu'ils définissent la hauteur du graphique alors que pour les doubles-masses la droite de la moyenne définissait la hauteur du graphique.

On regardera de nouveau:

- . si la courbe coupe souvent l'axe $y = 0$, il est alors probable qu'il n'y a pas de tendances significatives
- . si des grandes excursions d'un seul côté de l'axe $y = 0$ existent, une tendance est probable dans ce cas
- . si la forme est parabolique, une tendance linéaire monotone est suggérée
- . si on a des segments de droites discontinus, des tendances par saut sont probables.

6.1 Organigramme du programme P2.EXE



6.2 Contenu détaillé du programme P2.EXE

ACTION	RESTRICTIONS	SOUS-ROUTINES UTILISÉES
Étude des valeurs aberrantes	Les parties A) et B) sont effectuées jusqu'à ce que l'utilisateur ne rejette plus de valeurs aberrantes	
----- A) Paramètre mesuré vs Temps -----		GRAPH -----
B) Rejet du max et/ou du min comme valeur aberrante	- Choix de l'utilisateur	
Étude par double-masse Choix du menu 1- Concentration vs Temps 2- Charge vs Temps 3- Concentration vs Débit 4- Charge vs Débit	- Les choix 2,3,4, ne sont pas accessibles si les débits sont absents et que l'on traite les concentrations - Alors que les choix 1,3,4, ne le sont pas si l'on traite les charges (débits absents)	DOUBMASS
Étude par CUSUM Choix du menu 5- Débit vs Temps 6- Concentration vs Temps 7- Charge vs Temps	- Les choix 5,7 ne sont pas accessibles si les débits sont absents et que l'on traite les concentrations - Alors que les choix 5,6 ne le sont pas si l'on traite les charges (débits absents)	CUSUM
Étude complète Choix du menu: 8	- Non accessible si les débits sont absents	CUSUM DOUBMASS
Sortie Choix du menu: 9		

6.3 Figures du programme

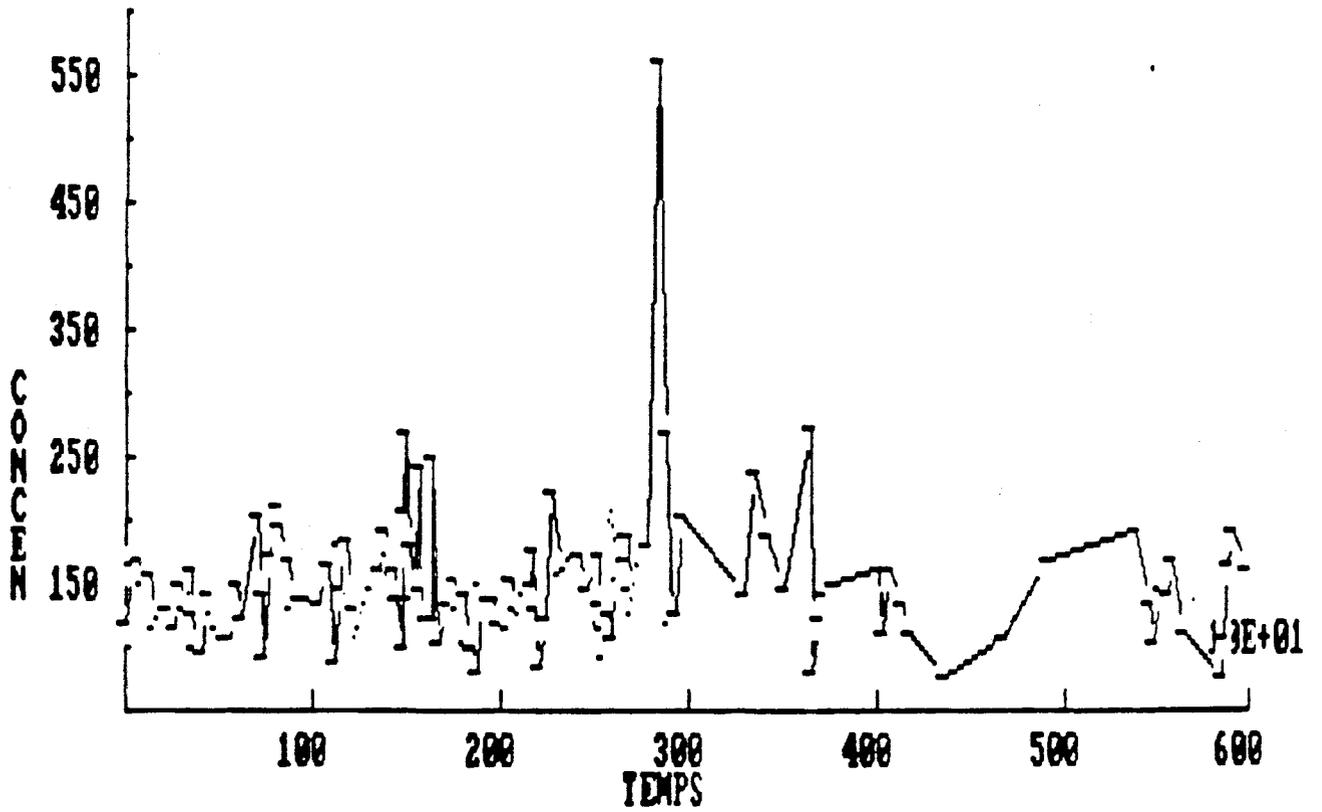
Le programme peut présenter jusqu'à 8 figures différentes dont une seule est présentée obligatoirement. Il s'agit de la figure 2.1 présentant l'évolution temporelle du paramètre et permettant une appréciation visuelle de la présence possible de valeurs aberrantes. Cette figure sera représentée après chaque élimination de valeur aberrante par l'utilisateur.

Les figures 2.2 à 2.5 présentent les différentes courbes doubles-masses appelées par le menu par les options 1, 2, 3 et 4. On cherche alors à détecter de grandes excursions d'un côté ou l'autre de la droite représentant la moyenne (droite rejoignant les coins inférieur gauche et supérieur droit), ces grandes excursions étant un indice important de la présence de tendances.

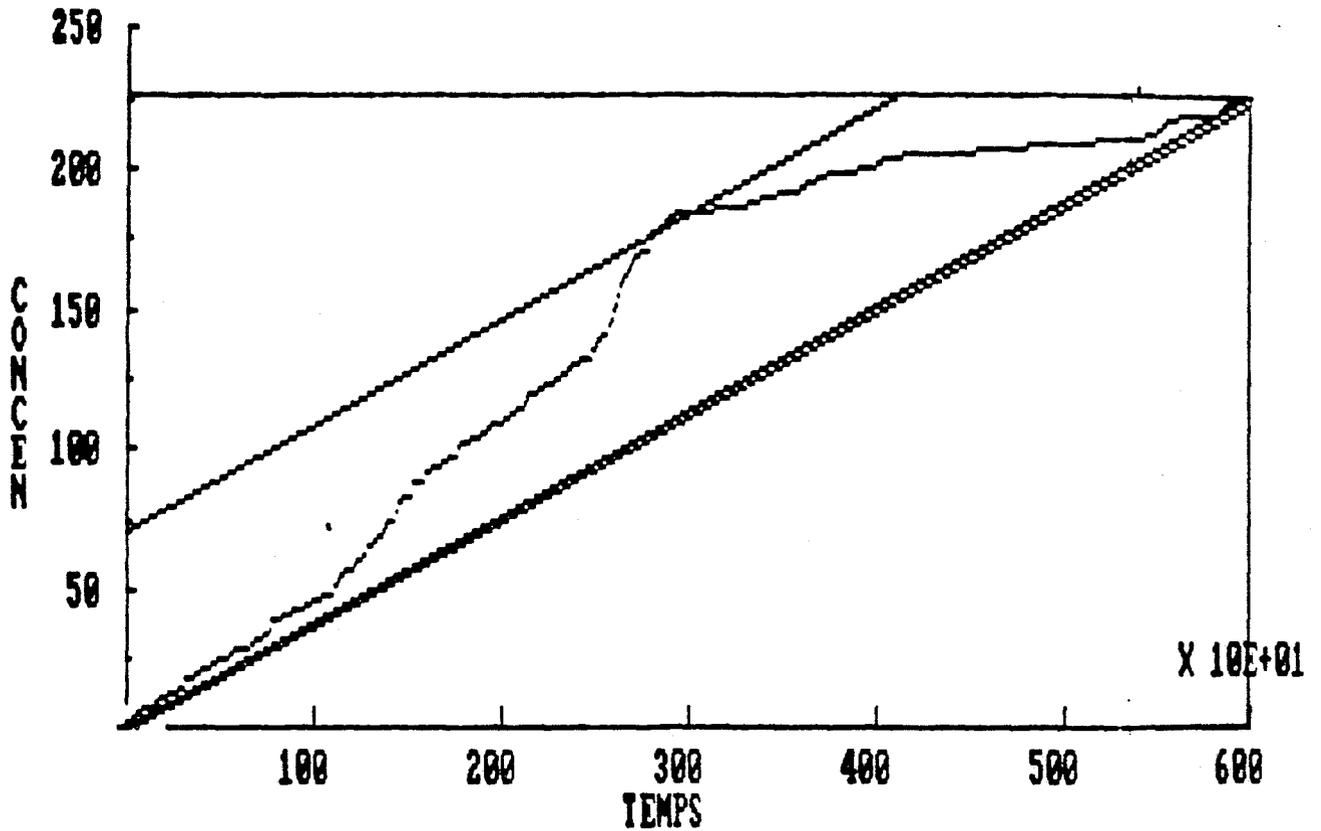
Les figures 2.6 à 2.8 présentent les courbes CUSUM permettant de visualiser sous forme de tendance, l'évolution temporelle de la concentration et la charge du paramètre ainsi que le débit qui leur est associé. Ces figures sont appelées par le menu et correspondent aux options 5,6 et 7.

Il faut noter que lorsque les débits ne sont pas disponibles, plusieurs figures ne sont pas accessibles: premièrement, si la série mesurée contient les concentrations et que les débits ne sont pas disponibles, les figures 2.3, 2.4, 2.5, 2.6 et 2.8 ne sont pas accessibles alors que si ce sont les charges qui sont mesurées, les figures 2.2, 2.4, 2.5, 2.6 et 2.7 ne sont pas accessibles.

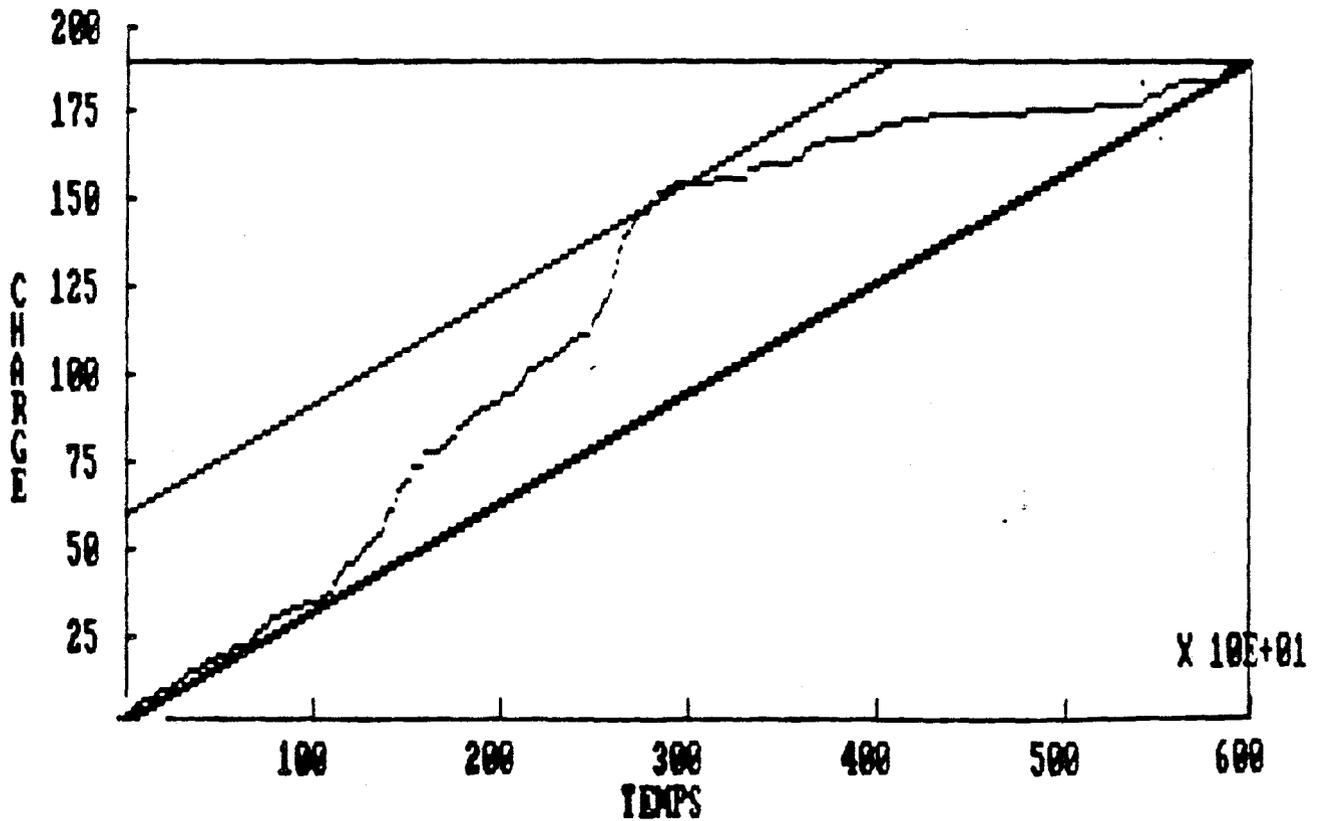
FIGURE 2.1 : CONCENTRATION VS TEMPS, FICH=9816CON



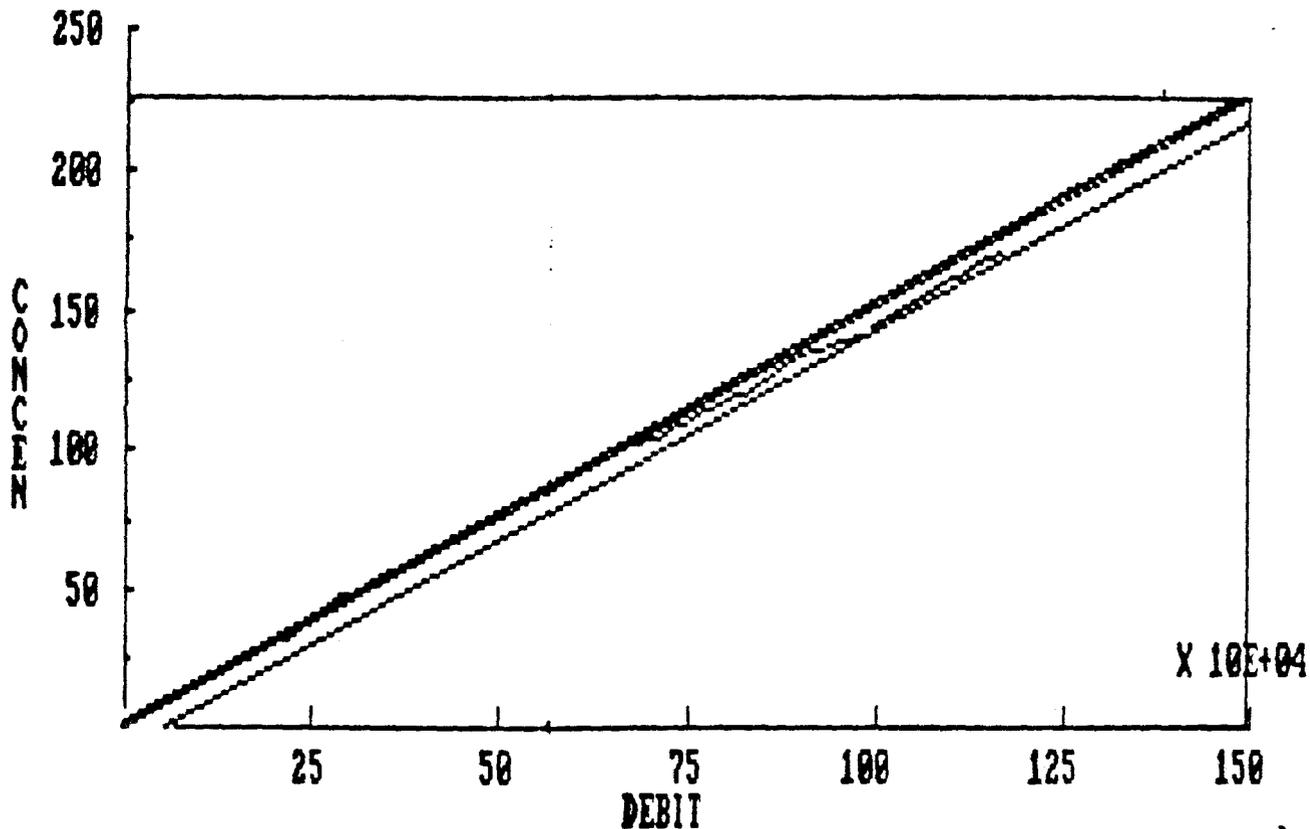
X 10E+02 FIGURE 2.2 :FONCTION DOUBLE-MASSE :C vs T, 9016CON



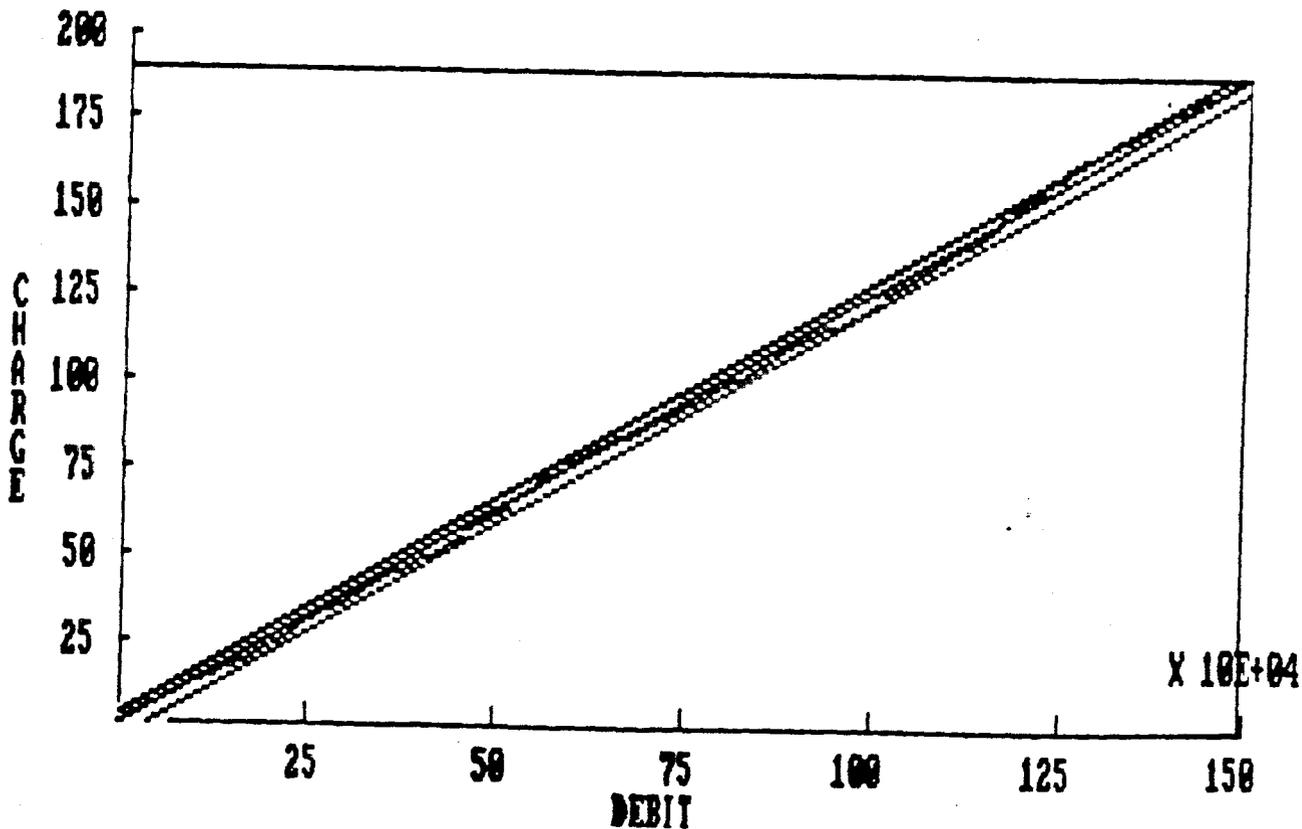
X 10E+11 FIGURE 2.3 :FONCTION DOUBLE-MASSE :L vs T, 9016CON



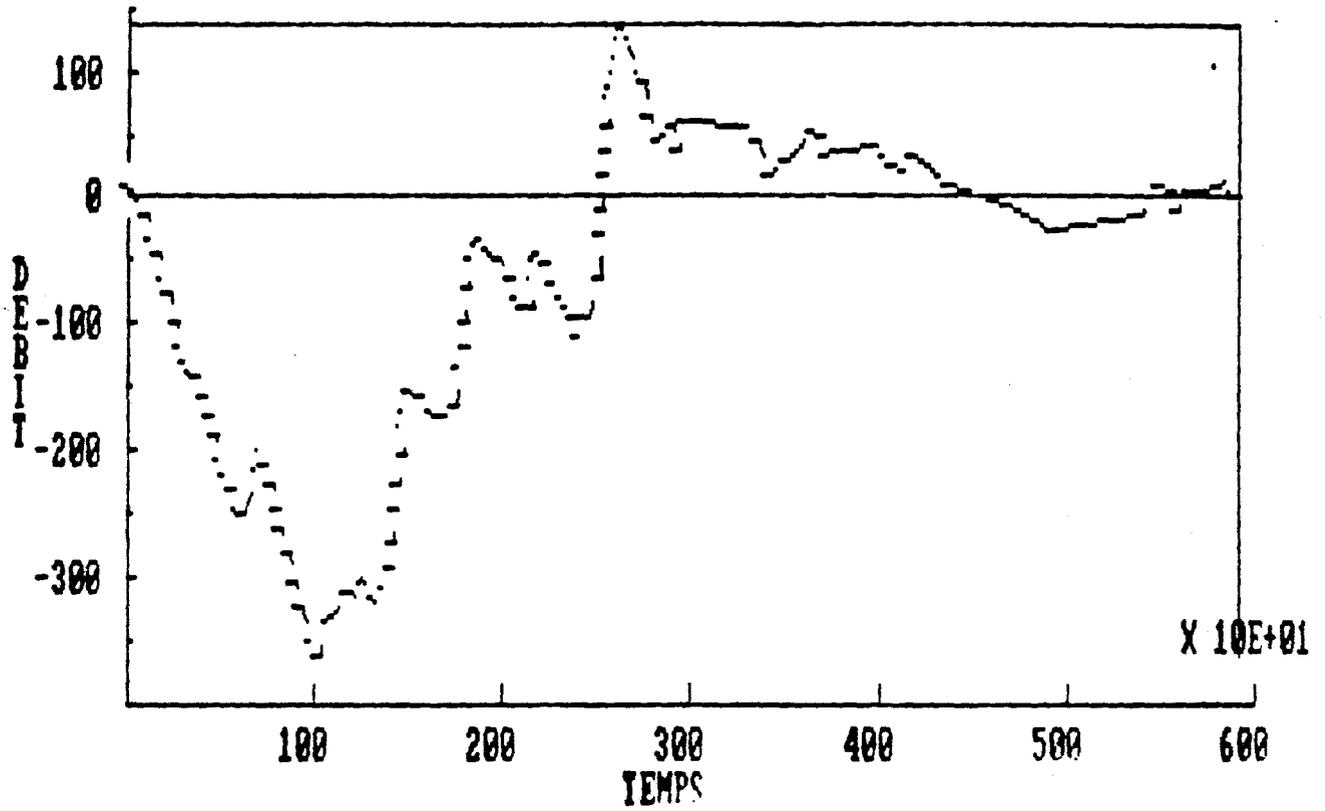
X 10E+02 FIGURE 2.4 : FONCTION DOUBLE-MASSSE : C vs Q, 9016CON



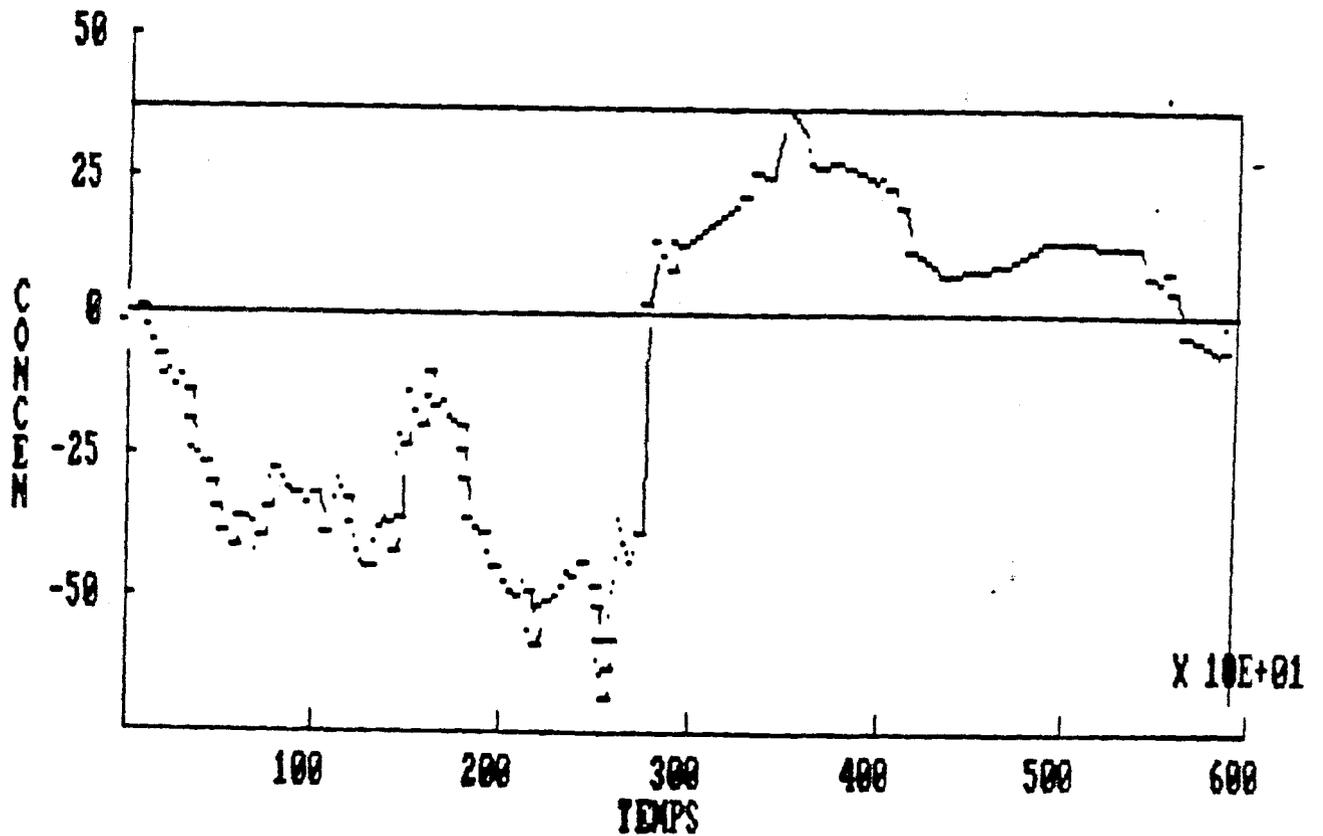
X 10E+11 FIGURE 2.5 : FONCTION DOUBLE-MASSSE : L vs Q, 9016CON



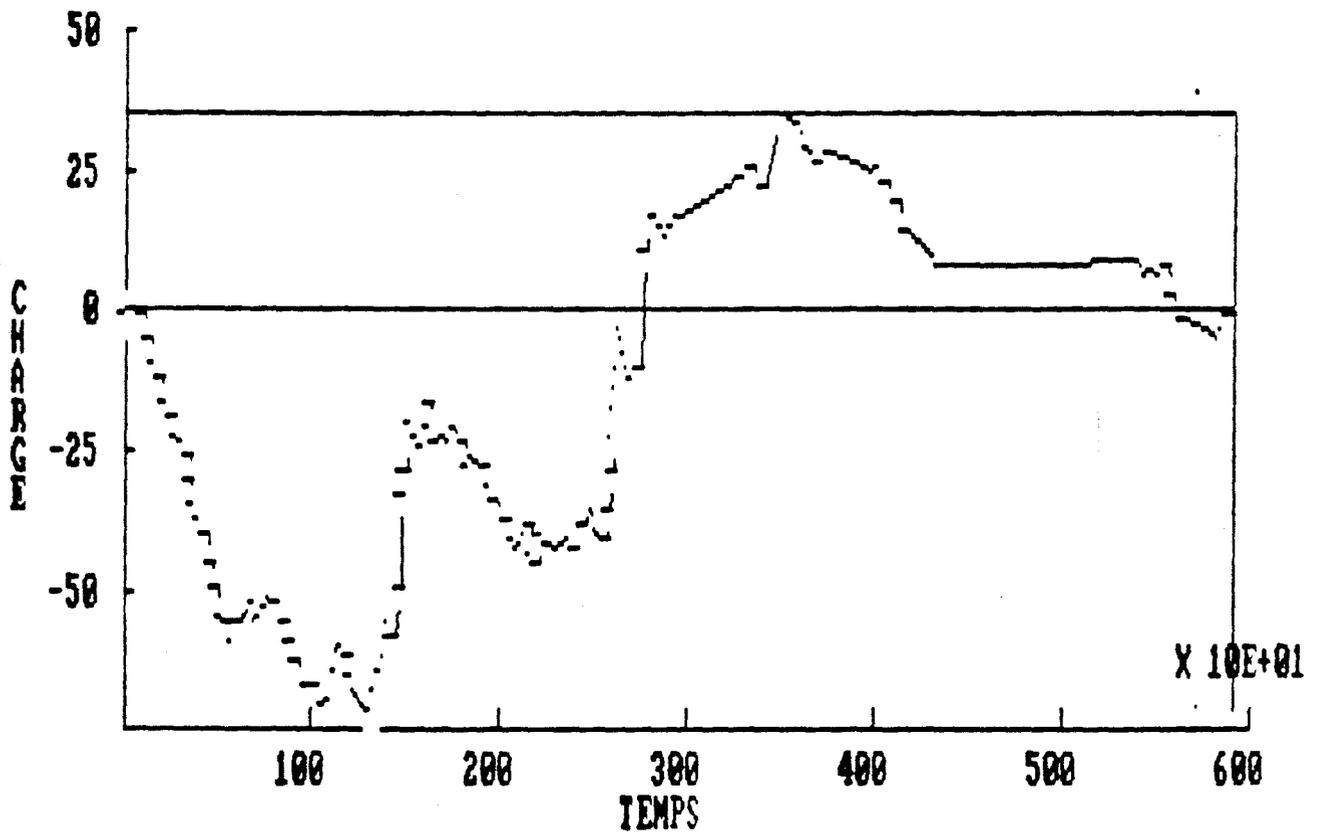
X 10E+02 FIGURE 2.6 : FONCTION CUSUM :Q vs T, 9816CON



X 10E+01 FIGURE 2.7 : FONCTION CUSUM :C vs T, 9816CON



X 10E+10 FIGURE 2.8 : FONCTION CUSUM :L vs T, 9816CON



7. Caractérisation de la série des mesures (P3.EXE)

Ce programme effectue les tâches suivantes:

- . présentation de l'évolution temporelle;
- . établissement d'un tableau des fréquences mensuelles d'échantillonnage. Ceci permet une première évaluation du pas de travail. Une question ultérieure permet d'éliminer des mois entiers où l'échantillonnage n'a pas lieu (par exemple, en hiver);
- . étude des saisonnalités sur une base mensuelle (par ANOVA). Par la suite le graphique des moyennes mensuelles suggère un regroupement des mois;
- . enfin recherche d'une relation concentration-débit significative qui pourra éventuellement être utilisée par la suite au choix de l'utilisateur.

GÉNÉRALITÉS

ÉCHANTILLONNAGE IRRÉGULIER

Après le tableau 3.1, certaines questions peuvent être posées en présence d'échantillonnage irrégulier.

- 1° Si au moins deux mois consécutifs n'ont jamais été échantillonnés. Il est alors offert à l'utilisateur de tronquer certains mois. La troncation doit alors être effectuée sur des mois consécutifs. Il n'est pas possible d'éliminer des mois non adjacents.

Lorsque des mois sont éliminés, le seul choix d'intervalles qui gardera cette troncation est celui de 1 valeur par mois pour les mois non-troncqués. Toute autre sélection ne tiendra pas compte de la troncation précédente.

N.B. Lors des utilisations pratiques du logiciel, il fut évident que le critère devait être un peu moins restrictif. Ainsi on offrira également de tronquer des mois si on a deux mois consécutifs avec deux valeurs ou moins.

2° Si au moins une année a moins de 4 valeurs. Il est alors offert à l'utilisateur de limiter le traitement ultérieur. Les choix possibles sont présentés dans le menu suivant:

Comme le nombre de valeurs pour certaine(s) année(s) est petit vous avez la possibilité de limiter le traitement.

- 1) Éliminer certaines années intermédiaires
- 2) Éliminer certains mois aux extrémités
- 3) Retour aux décomptes avant élimination
- 4) Voir le tableau complet actuel
- 5) Distribution non-représentative; sortie sans test
- 6) Suite normale du logiciel
- ?) Pour avoir de l'aide

Voici une brève description des différentes options de ce menu:

- 1) Éliminer certaines années intermédiaires

Il est alors offert à l'utilisateur de sélectionner pour traitement ultérieur, certaines années intermédiaires à faible fréquence. Les années retenues pour ce traitement doivent être consécutives. L'utilisateur aura le choix (dans P4) de:

- a) mettre les moyennes saisonnières dans les intervalles des années à faibles fréquences sélectionnées
- b) tronquer les années sélectionnées
- c) ne rien faire.

Lorsque b) est choisi, une étude de niveau (par saut) avant et après la troncation est choisie automatiquement par le logiciel en utilisant le test le plus approprié.

2) Éliminer certains mois aux extrémités

Il est alors offert à l'utilisateur de choisir une période de traitement plus petite que la période définie par le fichier d'entrée. Les valeurs ne faisant pas partie de la nouvelle période définie seront alors éliminées du fichier créé par P1(.TMP). Le reste du logiciel traitera donc cette série comme si les valeurs extérieures à cette nouvelle période n'existaient pas.

3) Retour aux décomptes avant élimination

Cette option permet de revenir à la période initiale si la nouvelle période définie n'est plus retenue par l'utilisateur.

4) Voir le tableau complet actuel

Cette option permet de revoir le tableau original. Ceci est utile pour bien définir la nouvelle période.

5) Distribution non-représentative: sortie sans test

Cette option permet de ne pas exécuter le reste du logiciel lorsque l'échantillonnage est trop faible ou trop irrégulier. On aura alors une sortie dans le fichier SYNTHESE.P6 qui contiendra les premières informations sur la série ainsi que le tableau 3.1 présentant les fréquences mensuelles. Il sera question dans le rapport d'applications de quelques critères permettant de détecter un échantillonnage trop faible ou trop irrégulier

6) Suite normale du logiciel

Cette option permet de retourner à la suite normale du logiciel une fois les traitements effectués. On peut également décider de ne pas traiter les années à faible fréquence et d'utiliser cette option dès la première apparition du menu.

?) Pour avoir de l'aide

Donne quelques informations sur l'utilisation de l'option (1).

ANOVA

Les analyses de variance effectuées dans le logiciel n'utilisent qu'une seule valeur par année par "saison". Si on a plus d'une observation avec l'échantillonnage original, la moyenne de ces observations est alors utilisée. Une telle manipulation permet de traiter des valeurs qui devraient être indépendantes puisque si l'on a 10 observations par mois et qu'il est probable que ces observations soient persistantes, on viole alors l'hypothèse d'indépendance des observations reliée à l'analyse de variance.

Il faut cependant noter que si les tailles d'échantillons par intervalle sont très différentes, l'utilisation de la moyenne peut causer certains ennuis. Il faudra alors être critique des résultats obtenus et l'utilisation de tests saisonniers et non-saisonniers pourra permettre de valider les conclusions. Comme aucune planification d'expériences n'a été faite afin d'avoir des résultats optimaux, on doit donc se servir de l'analyse de variance avec les données disponibles et valider les conclusions lorsqu'elle ne semble pas adéquate.

Il faut cependant noter que l'analyse de variance utilisée ici est une étape secondaire à la détection de tendance puisqu'elle ne sert qu'à savoir s'il y a saisonnalité.

Afin de valider l'utilisation de l'analyse de variance, un test sur l'égalité des variances (BARTLETT) est effectué avant l'impression des résultats et un avertissement est donné à l'utilisateur dans le cas d'un rejet de l'égalité des variances.

SAISONNALITÉ

L'utilisation des analyses de variance a pour but de construire des "saisons" ayant des moyennes significativement différentes. On effectue d'abord une analyse de variance pour tester l'égalité des moyennes mensuelles. Si les moyennes sont testées comme différentes, on accepte la présence de saisonnalité. On offre alors à l'utilisateur d'effectuer des regroupements qui permettront de traiter des saisons plus grandes (cette possibilité de regroupement n'est pas possible si on a eu troncation de certains mois). L'utilisateur pourra alors effectuer autant de regroupements qu'il le désire, seul le dernier regroupement accepté sera gardé en mémoire. Si aucun regroupement n'est accepté par l'utilisateur, on gardera en mémoire une saisonnalité formée de 12 saisons de 1 mois.

L'utilisation de saisonnalités permet d'effectuer des tests qui tiendront compte de la présence de cycles dans les données. Cependant, l'utilisation de tests saisonniers, sur des données ne contenant pas de saisonnalités, amènera une perte de puissance par rapport à un test non-saisonnier.

RELATION C-Q

La recherche d'une relation fortement significative entre la concentration et le débit a pour but de permettre une méthode alternative pour estimer de façon adéquate une valeur pour les intervalles vides.

La relation recherchée est de la forme:

$$C = aQ^b$$

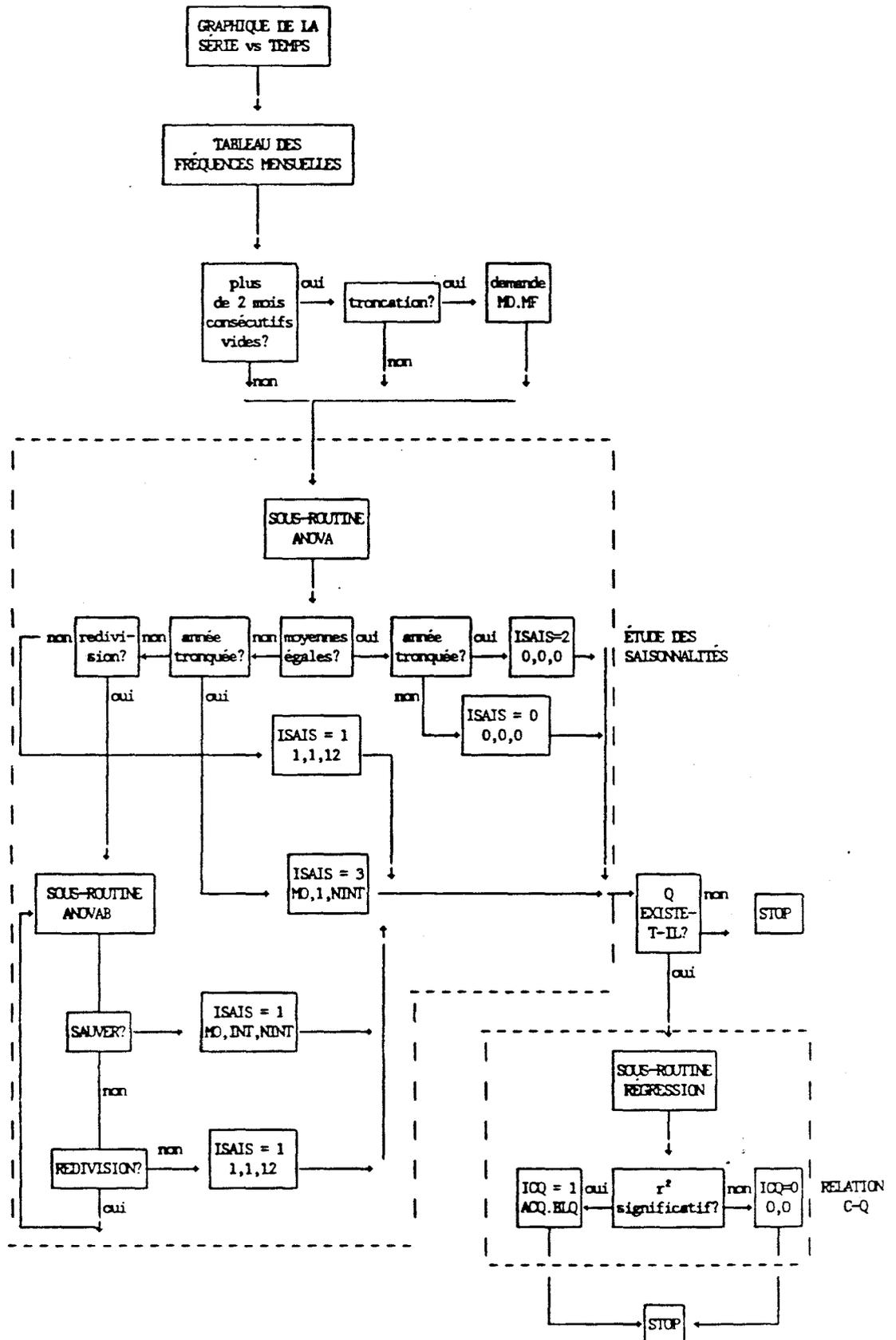
on effectue d'abord la régression pour:

$$\ln C = a^* + b^* \ln Q$$

et on revient aux unités initiales avec $\hat{a} = e^{a^*}$ et $b = b^*$. Ce modèle n'est pas le meilleur modèle pour la base originale cependant une telle construction est simple et nous permet d'avoir quand même une bonne idée de la force de la relation entre la concentration et le débit.

Afin de conclure à une relation significative, le logiciel effectue le test associé à FISHER sur les coefficients de corrélation. On demande de plus que le pourcentage de variance expliquée soit plus grand que 25% donc un coefficient de corrélation plus grand que .50.

7.1 Organigramme du programme P3.EXE



7.2 Contenu détaillé du programme P3.EXE

ACTION	RESTRICTIONS	SOUS-ROUTINES UTILISÉES
Paramètre vs Temps (RAN6)	Seule la série choisie est présentée, concentration ou charge ref. P1	GRAPHI
Tableau des fréquences mensuelles		FREQUENS
Troncation de l'année	Seulement si deux mois consécutifs n'ont aucun échantillon pour toutes les années documentées	
Étude de saisonnalités A) Analyse de variance avec 12 saisons de 1 mois		ANOVA
B) Graphique des moyennes mensuelles		MOYMENS GRAF
C) Analyse de variance avec regroupement de mois	- Seulement si les moyennes mensuelles sont testées différentes dans A) - Jamais si on a eu troncation	ANOVAB
Relation concentration-débit	- Jamais si les débits ne sont pas présents	REGRESSION

7.3 Figures et tableaux du programme

Le programme contient 5 figures et 6 tableaux. Cependant, ils ne seront jamais tous présentés lors d'une même utilisation puisque dans certains cas on a une figure ou un tableau différent pour la charge et la concentration.

A) Figures

Les figures 3.1 et 3.2 présentent l'évolution temporelle du paramètre étudié, une seule de ces figures sera présentée selon le choix de l'utilisateur de traiter la charge ou la concentration.

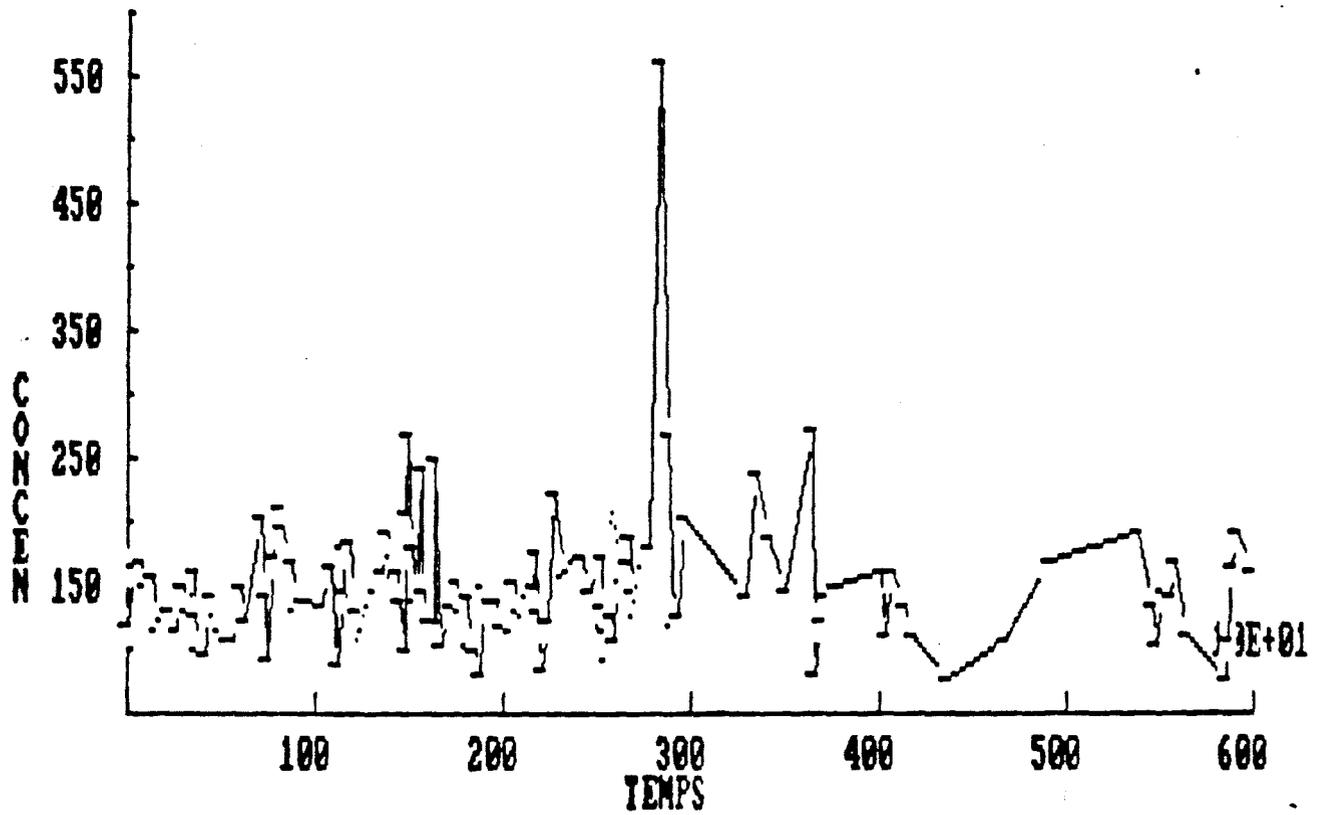
Les figures 3.3 et 3.4 présentent les moyennes mensuelles de chaque année du paramètre étudié. Les symboles utilisés (1-9) représentant le dernier chiffre de l'année alors que les étoiles reliées par des segments de droite représentent les moyennes de toutes les valeurs d'un même mois. Ce graphique permet une bonne visualisation des regroupements possibles de mois sous forme de saisons homogènes.

La figure 3.5 présente les logarithmes des concentrations en fonction des logarithmes des débits permettant ainsi de visualiser la relation pouvant être présente entre les deux variables et complètent l'analyse de régression présentée dans le tableau 3.4.

B) Tableaux

Le tableau 3.1 présente les fréquences mensuelles d'observation pour chaque année. On peut alors avoir une bonne idée de l'intervalle à choisir pour créer la série équidistante (P4.EXE). Lors d'un échantillonnage suffisamment régulier, le nombre moyen d'observations par an devient un choix de fréquence intéressant qui ne devrait pas donner lieu à trop d'intervalles vides.

FIGURE 3.1 : CONCENTRATION VS TEMPS, FICH=9816CON



X 10E+09 FIGURE 3.2 : CHARGE VS TEMPS, FICH=9816CON

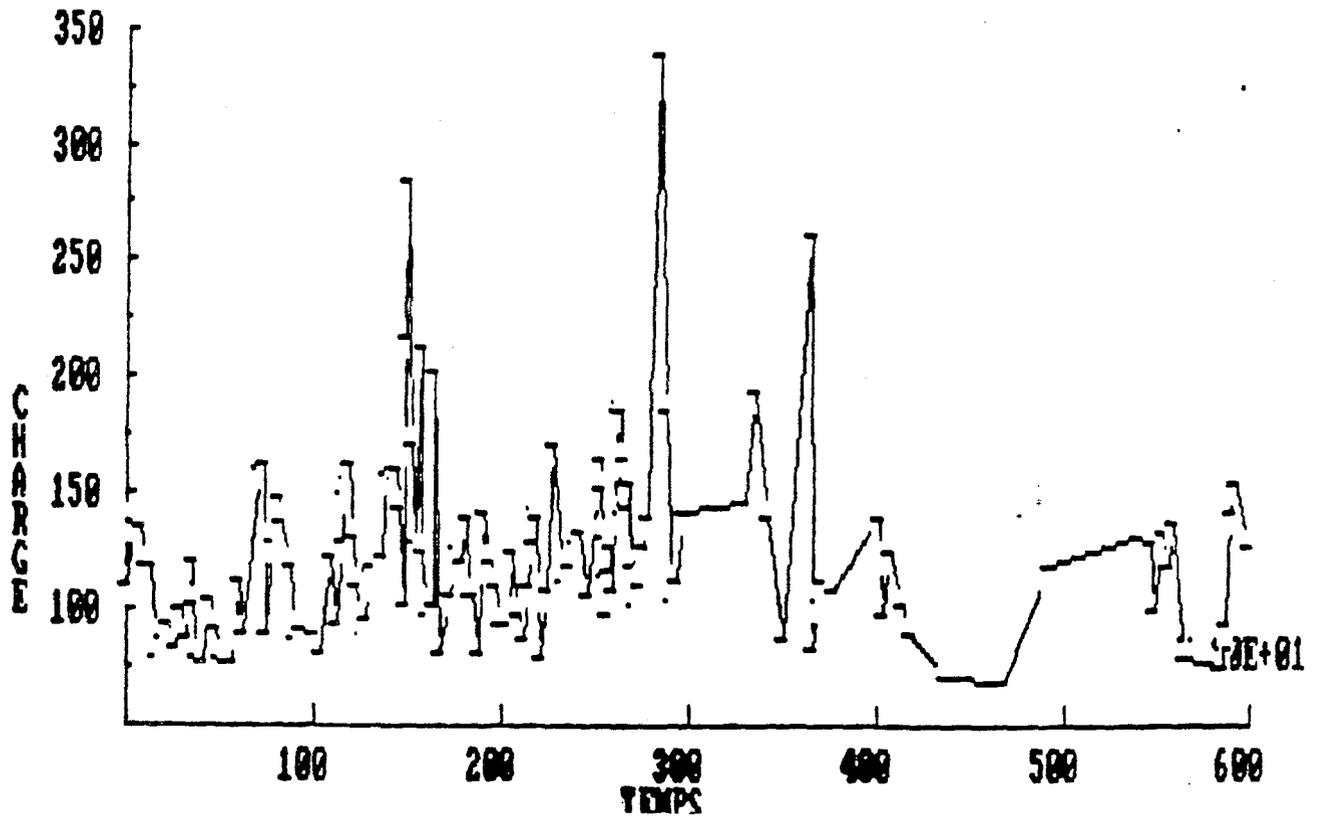
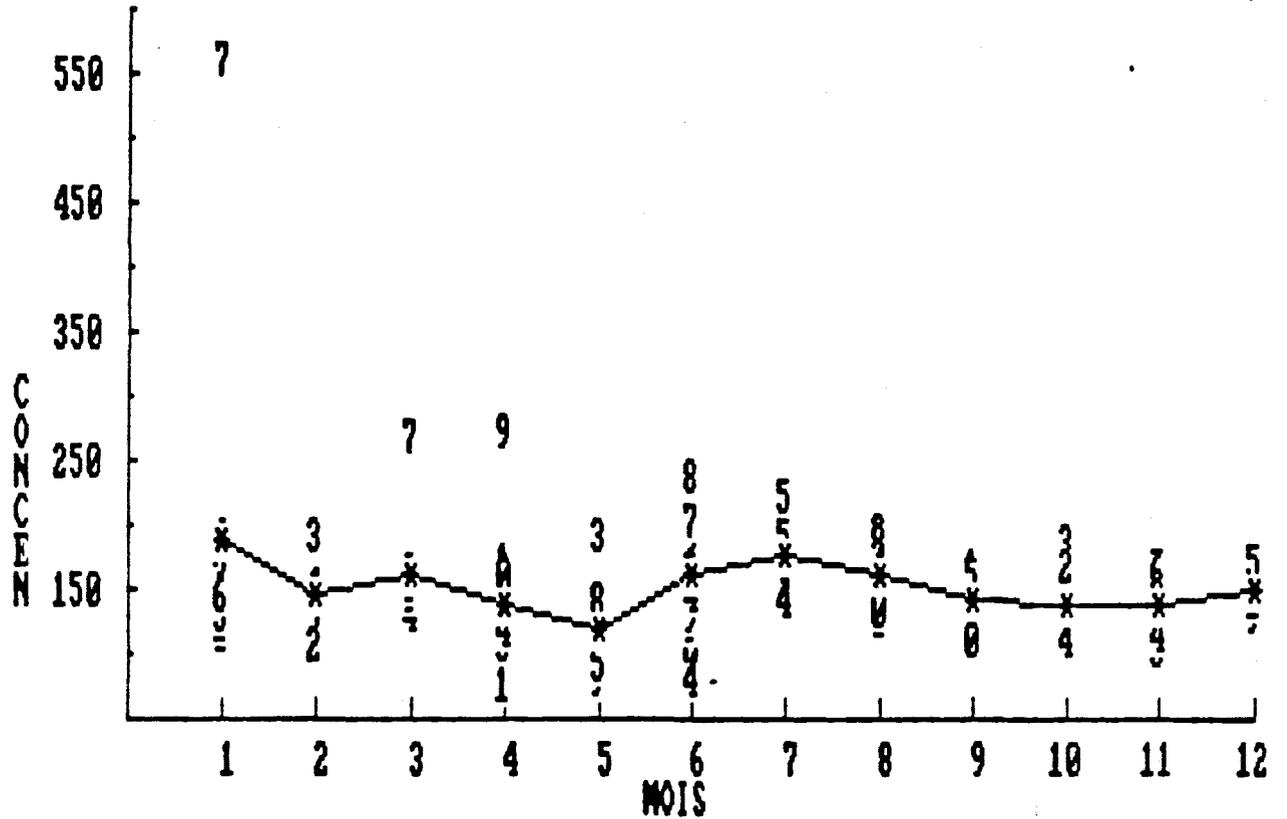
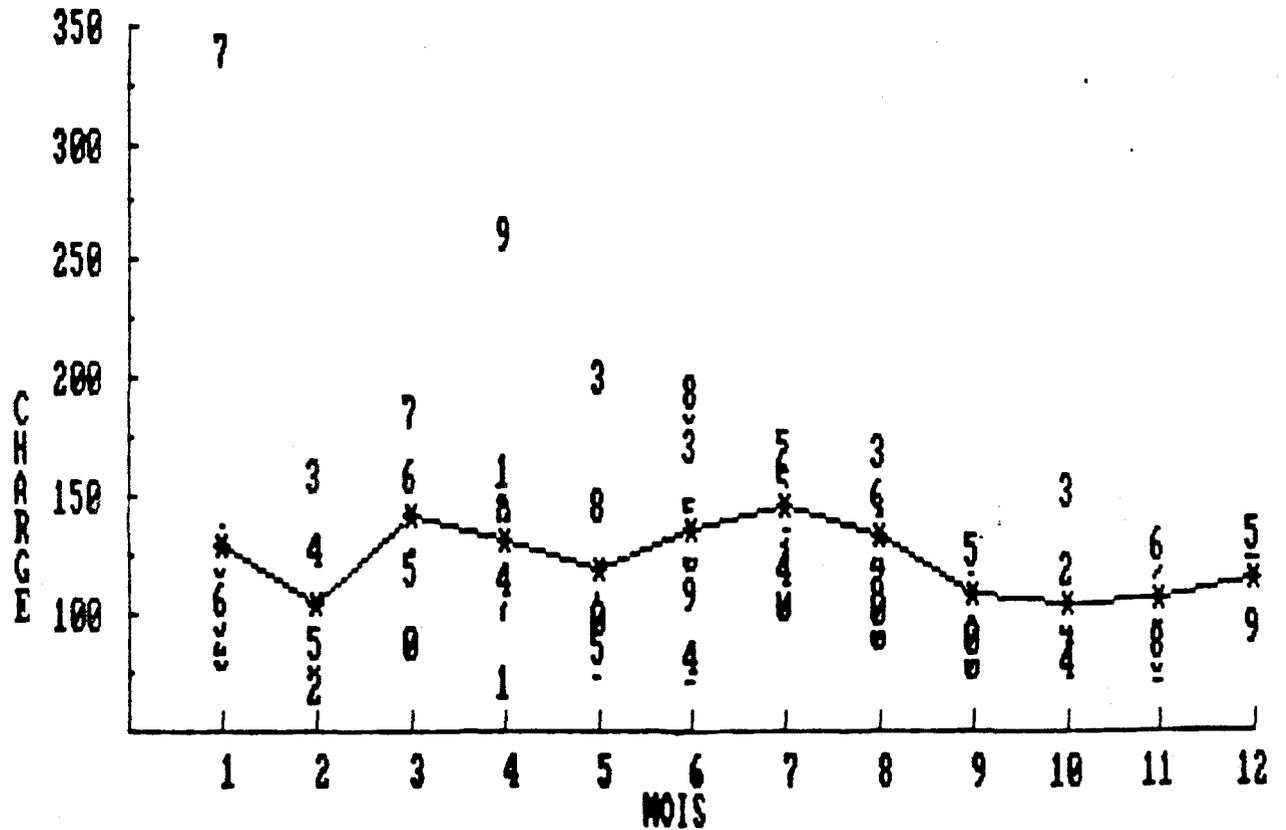


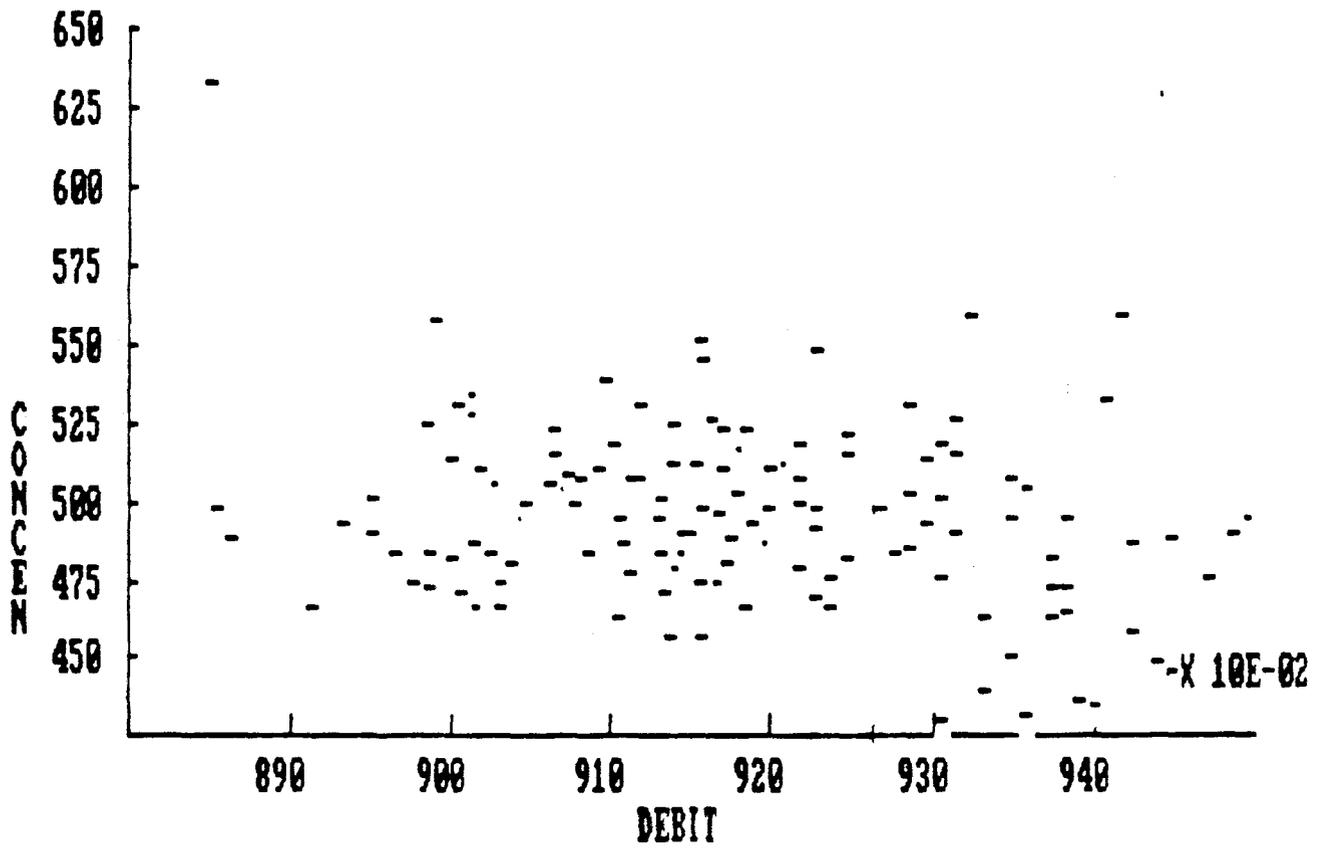
FIGURE 3.3: MOYENNES MENSUELLES DES CONCENTRATIONS, FICH=9016CON



X 10E+09 FIGURE 3.4: MOYENNES MENSUELLES DES CHARGES, FICH=9016CON



X 10E-02 FIGURE 3.5 : Ln C VS Ln Q, FICH=9016CON



Le tableau 3.2 présente les résultats de l'analyse de variance sur l'égalité des moyennes mensuelles. Chaque observation de la série originale représente alors une répétition pour le mois durant lequel elle a été mesurée. Les résultats sont présentés selon la forme habituelle des tables ANOVA, ainsi:

- d.l. = nombre de degré de liberté
- ss = somme des carrés (sum of square)
- ms = moyenne des carrés (mean square)
- F = valeur de la statistique pour le test $H_0: \mu_1 = \mu_2 = \dots = \mu_{12}$ vs H_1 : les moyennes mensuelles ne sont pas égales, où μ_i = moyenne du i^e mois.

L'utilisateur intéressé à en savoir plus long sur l'analyse de variance et les hypothèses sous-jacentes à son utilisation peut consulter NETER et WASSERMAN (1974): Applied Linear Statistical Models. Irwin. Homewood. 842 p.

On acceptera la présence de saisonnalité si les moyennes mensuelles sont testées comme différentes, c'est-à-dire si F est plus grand que la valeur théorique de la loi de FISHER avec les degrés de liberté présents dans la table. Cette comparaison est faite automatiquement par le logiciel au seuil $\alpha = 5\%$ et le résultat est présenté dans la partie commentaires. Le tableau 3.3 est présenté plutôt que le tableau 3.2 si l'on traite la charge. Lorsque l'on rejette l'égalité des moyennes, l'utilisateur a la possibilité de regrouper des mois afin de construire des saisons lui apparaissant plus appropriées.

Le tableau 3.2b présente les résultats de l'analyse de variance effectuée à partir du regroupement proposé par l'utilisateur. Cette table ANOVA présente les mêmes caractéristiques que celle du tableau 3.2 sauf pour le test correspondant qui vérifie alors $H_0: \mu_1 = \dots = \mu_k$ où k est le nombre de saisons définies par le regroupement. Le tableau 3.3b est présenté si la charge est traitée plutôt que la concentration.

Le tableau 3.4 présente les résultats de la régression sous forme de la relation $C = aQ^b$. Les premiers résultats: variance et moyenne, concernent les logarithmes naturels des concentrations et des débits. Les estimateurs des paramètres a et b sont obtenus à partir du résultat de la régression $\ln C = a^* + b^* \ln Q$, en transformant $a = e^{a^*}$ et $b = b^*$. Le pourcentage de variance expliqué permet une évaluation de la force de la relation estimée. Le test associé à FISHER est également effectué afin de savoir si la relation est significative et permet une estimation valable des concentrations à partir des débits.

Tableau 3.1: Fréquences mensuelles observées

	J	F	M	A	M	J	J	A	S	O	N	D	TOTAL/AN
1969	0	0	0	0	0	1	1	1	1	1	1	1	7
1970	1	1	1	2	1	1	1	1	1	1	1	1	13
1971	1	0	0	2	1	1	2	0	1	1	1	0	10
1972	1	1	1	1	1	1	1	1	0	1	1	1	11
1973	1	1	2	1	1	1	1	1	1	1	1	0	12
1974	1	1	1	3	1	1	1	1	1	1	1	0	13
1975	1	1	2	2	1	1	1	0	1	1	1	1	13
1976	1	0	2	3	3	2	4	4	3	2	3	0	27
1977	1	0	1	1	1	1	0	0	0	0	0	0	5
TOT/MOIS	8	5	10	15	10	10	12	9	9	9	10	4	111
	Nombre moyen d'observation/an:												12.333
	Écart-type du nombre de mesures annuelles:												5.831

Tableau 3.2: Anova pour les concentrations

TABLE ANOVA POUR L'ÉGALITÉ DES MOYENNES MENSUELLES

Source	d.l.	ss	ms	F
Mois	11	.39600E+05	.36000E+04	4.707
Erreur	90	.75718E+05	.76483E+03	

Total	110	.11532E+06	.10483E+04	
-------	-----	------------	------------	--

Commentaires: on rejette l'égalité des moyennes

Tableau 3.2b: Anova pour les concentrations

TABLE ANOVA POUR L'ÉGALITÉ DES MOYENNES MENSUELLES

Source	d.l.	ss	ms	F
Mois	3	.14111E+05	.47036E+04	4.973
Erreur	107	.10121E+06	.94586E+03	

Total	110	.11532E+06	.10483E+04	
-------	-----	------------	------------	--

Commentaires: on rejette l'égalité des moyennes

Tableau 3.3: Anova pour les charges

TABLE ANOVA POUR L'ÉGALITÉ DES MOYENNES MENSUELLES

Source	d.l.	ss	ms	F
Mois	11	.21273E+19	.19339E+18	2.700
Erreur	99	.70924E+19	.71640E+17	

Total	110	.92197E+19	.83816E+17	
-------	-----	------------	------------	--

Commentaires: on rejette l'égalité des moyennes

Tableau 3.3b: Anova pour les charges

TABLE ANOVA POUR L'ÉGALITÉ DES MOYENNES MENSUELLES

Source	d.l.	ss	ms	F
Mois	3	.10975E+19	.36585E+18	4.820
Erreur	107	.81222E+19	.75908E+17	

Total	110	.92198E+19	.83816E+17	
-------	-----	------------	------------	--

Commentaires: on rejette l'égalité des moyennes

Tableau 3.4: Résultats de la régression

Moyenne des LN(CONC):	4.94
Variance:	.05
Moyenne des LN(DEB) :	4.26
Variance:	.04

Relation $C=a Q^{**b}$

a = 121.11 b= .03
% de variance expliquée : 1.29

8. Série de travail, création et structure (P4.EXE)

Cette section du logiciel permet d'effectuer les tâches suivantes:

- . choix d'un intervalle de travail
 - jours (1 3 7 15)
 - mois (1 2 3 4 6 12)

- . régénérer des valeurs absentes selon 3 méthodes:
 - a) interpolation temporelle
 - b) moyenne saisonnière
 - c) relation concentration-débit

À ce stade, on a généré une série complète de valeurs équidistantes, sauvées dans un fichier avec l'extension .TMC ou .TML selon qu'il s'agit de concentrations ou de charges.

La dernière partie évalue la persistance de cette série et, si elle est significative, détermine si elle peut être markovienne du 1er ordre.

GÉNÉRALITÉS

INFORMATION FICTIVE

Le choix d'un intervalle de travail trop fin amènera la création d'information fictive. Il est important de ne pas combler un trop grand nombre d'intervalles à l'aide des méthodes proposées (moyennes saisonnières, interpolation, relation concentration-débit). Afin de mettre en garde l'utilisateur d'une telle éventualité, un message est écrit lorsque plus de 20% des intervalles devront être générés à l'aide des autres données. L'utilisateur aura alors la possibilité de choisir un autre intervalle de travail afin de réduire le nombre d'intervalles sans donnée, ou il pourra continuer, tout en sachant que sa série finale contient un bon pourcentage d'information fictive car, dans certains cas d'échantillonnage très irrégulier, il peut être difficile de faire mieux.

MÉTHODES POUR GÉNÉRER DES DONNÉES DANS DES INTERVALLES VIDES

Les trois méthodes offertes pour remplir les intervalles vides sont:

- l'interpolation
- utilisation de la moyenne de l'intervalle pour les autres années
- la relation concentration-débit

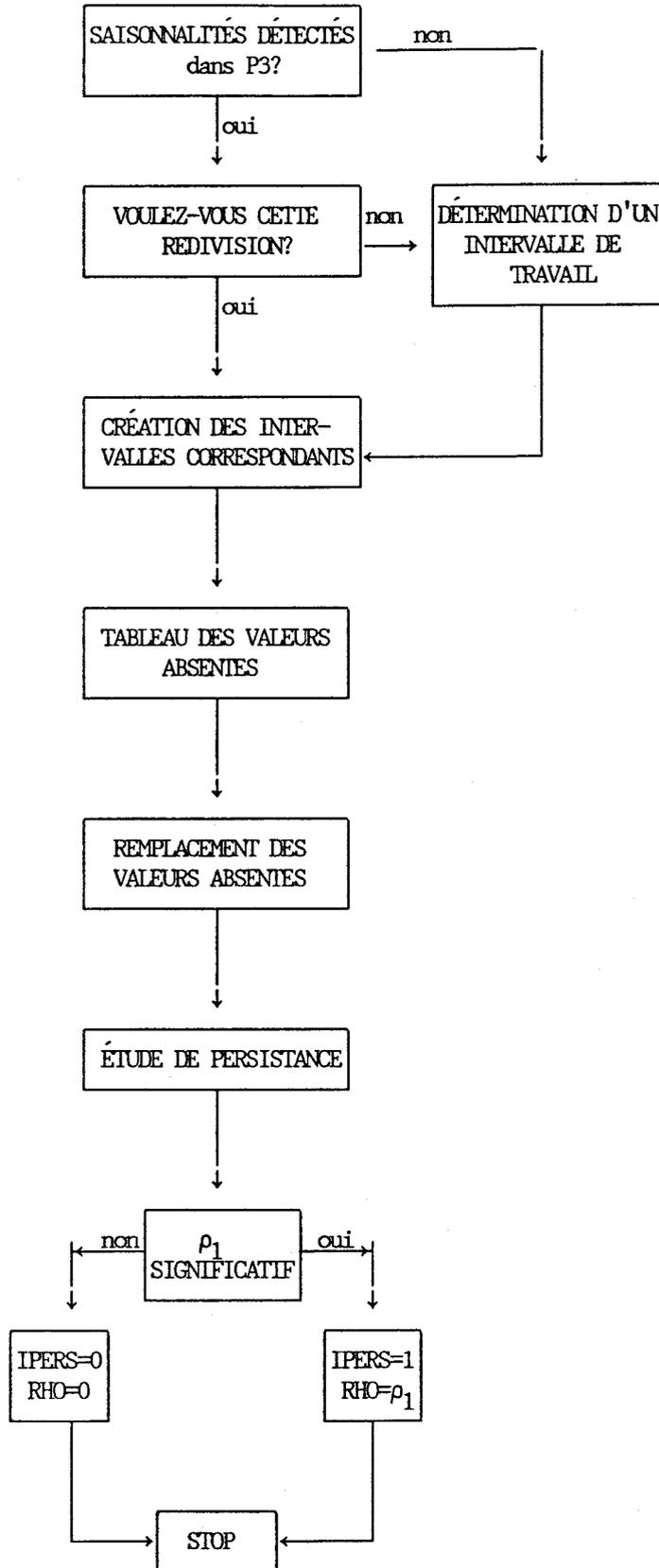
L'interpolation sera un choix intéressant lorsque l'on a pas de saisonnalités et que l'on a de la persistance. Il faut cependant prendre note que l'utilisation de l'interpolation augmentera cette persistance.

L'utilisation de la moyenne de l'intervalle sera surtout adéquate lorsque l'on a détecté une saisonnalité et qu'elle est utilisée, puisque l'on utilisera alors des moyennes représentatives de chacun des intervalles.

La relation concentration-débit serait un complément intéressant des deux autres méthodes mais il semble que ces relations ne soient pas souvent assez fortes pour permettre une estimation valable pour les intervalles vides.

N.B. Lorsque le nombre d'observations est petit, il peut être plus intéressant de choisir la méthode des moyennes afin de ne pas créer de persistance avant l'étude de cette même persistance. En effet, on peut noter ici une difficulté du traitement: on doit avoir des données équidistantes pour faire l'étude de persistance et il est préférable d'avoir l'étude de persistance pour choisir l'interpolation comme méthode de génération des données.

8.1 Organigramme du programme P4.EXE



8.2 Contenu détaillé du programme P4.EXE

ACTION	RESTRICTIONS	SOUS-ROUTINES UTILISÉES
Détermination d'un intervalle de travail	S'il y a eu saisonnalité détectée dans P3, il est possible de reprendre ces saisons.	ENSAISONS, TRONQUÉE ENJOURS, ENMOIS
Affichage des données manquantes par intervalle	Seulement s'il y a au moins une donnée manquante.	AFFICHE
Remplacement des données manquantes	Seulement s'il y a au moins une donnée manquante. Si on a décidé de ne pas utiliser les débits dans P1 ou si la relation concentrations-débits n'est pas significative (P3), cette dernière méthode n'est pas autorisée.	METMOV INTERPOL CONCDEB
Étude de persistance		AUTO, AUTOCO

8.3 Tableaux du programme

Le programme présente 2 tableaux. Le tableau 4.1 présente le nombre de valeurs absentes par intervalle: le nombre d'intervalles équivaut au nombre d'intervalles complets à l'intérieur d'une année. Les intervalles sont numérotés à partir du début de l'année civile s'il n'y a pas eu troncation. Le nombre de valeurs absentes pour un intervalle correspond au nombre d'années pour lesquelles aucune observation n'est disponible pour cet intervalle. Si le nombre de valeurs absentes est élevé pour l'ensemble des intervalles, on peut alors croire que l'intervalle choisi est trop fin (fréquence trop élevée). À l'opposé, un faible nombre de valeurs absentes peut être le résultat d'un bon choix d'intervalle dans le cas d'un échantillonnage régulier ou encore le choix d'un intervalle trop large (fréquence trop faible). Il faut noter que ce tableau n'apparaît pas, si aucune valeur n'est absente (aucun intervalle vide).

Le tableau 4.2 présente l'étude de persistance effectuée avec la série équidistante. On voit d'abord les coefficients d'autocorrélation d'ordre 1 à 6 et leurs écart-types. On peut alors avoir une idée des coefficients significativement différents de 0 et ainsi connaître la structure probable de la persistance. Un jugement est porté par le logiciel sur la structure la plus probable de persistance. Trois structures peuvent alors être proposées: premièrement, pas de persistance si le coefficient d'autocorrélation d'ordre 1 (ρ_1) n'est pas significativement différent de 0, deuxièmement, présence de persistance markovienne si ρ_1 est significativement différent de 0 et le coefficient d'autocorrélation partielle d'ordre 2 n'est pas significativement différent de 0. Finalement, on propose une persistance non-markovienne si ρ_1 et le coefficient d'autocorrélation partielle d'ordre 2 sont significativement différents de 0.

On considère qu'un coefficient est significativement différent de 0 si sa valeur est au moins 2 fois plus grande que son écart-type.

Tableau 4.1: Nombre de valeurs absentes par intervalle

NOMBRE D'ANNÉES CONSIDÉRÉES: 10

INTERVALLE	NOMBRE DE VALEURS ABSENTES	PROPORTION ABSENTES%
1	1	10.
2	4	40.
3	2	20.
4	1	10.
5	1	10.
6	0	0.
7	1	10.
8	3	30.
9	2	20.
10	1	10.
11	1	10.
12	5	50.

TOTAL D'INTERVALLES / # INTERVALLES OCCUPÉS : 1.22

Tableau 4.2: Étude de persistance

AUTOCORRÉLATION

	1	2	3	4	5	6
COEFFICIENT	.59	.30	.19	.09	.02	.04
ÉCART-TYPE	.10	.13	.14	.14	.14	.14

AUTOCORRÉLATION

	1	2
COEFFICIENT	.59	-.06
ÉCART-TYPE	.10	.10

Commentaire: Présence de persistance markovienne

9. Préparation aux tests (P5.EXE)

Ce programme rassemble toutes les caractéristiques de la série de travail mises en évidence précédemment: intervalle, saisonnalité, persistance, longueur, etc.

Il s'informe ensuite de la partie de série à traiter, de la recherche d'une tendance monotone ou par saut et suggère un test selon l'arbre de décision suivant:

ARBRE DE DÉCISION

<u>TENDANCE</u>	<u>PERSISTANCE</u>	<u>SAISONNALITÉ</u>	<u>TEST APPROPRIÉ</u>
Tendance → monotone	Persistance	- pas de saison	Lettenmaier/Spearman
	markovienne	- avec saisons	Hirsch et Slack
	Pas de	- pas de saison	Spearman/Kendall
	persistance	- avec saisons	Kendall saisonnier
Tendance → par saut	Persistance	- pas de saison	Lettenmaier/Mann-Whitney
	markovienne	- avec saisons	Hirsch et Slack
	Pas de	- pas de saison	Mann-Whitney
	persistance	- avec saisons	Kendall saisonnier

Le test choisi par l'utilisateur est effectué et ses résultats sont présentés.

GÉNÉRALITÉS

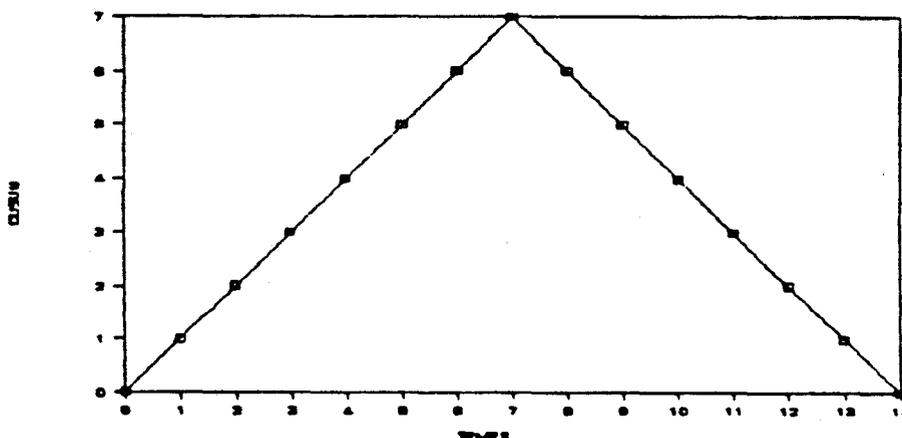
MONOTONE OU PAR SAUT?

Le choix de la détection d'une tendance monotone ou par saut peut être fait selon la connaissance de l'utilisateur du phénomène sous-jacent à la série de mesure. L'objectif de l'analyse aide aussi à un tel choix; par exemple si on veut tester l'impact de l'implantation d'une usine, il serait intéressant de choisir la détection d'une tendance par saut avec la séparation à la date de

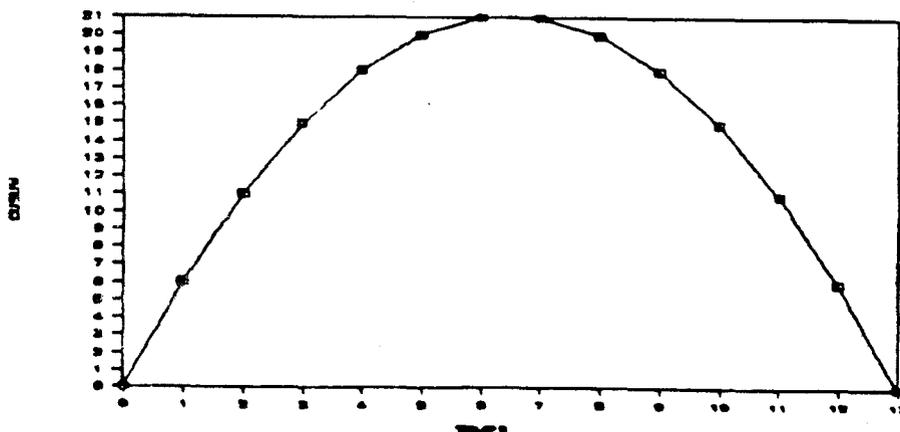
mise en marche de l'usine. Dans le cas où l'on veut voir un changement graduel comme l'effet des pluies acides, il peut être plus intéressant de choisir la détection d'une tendance monotone.

Lorsque la tendance est forte, les deux méthodes détecteront cette tendance et le critère d'ajustement présenté à la fin de P6 pourra permettre de choisir la méthode la plus adéquate.

Les fonctions CUSUM peuvent être très utiles d'un point de vue exploratoire pour faire un tel choix. En effet une tendance par saut amènera un CUSUM de la forme:



alors qu'une tendance monotone amènera un CUSUM de la forme:



DOUBLE-MASSE ET CUSUM

Ces graphiques doivent être analysés de la même façon que dans la partie P2. Il faut cependant prendre note que les données sont maintenant équidistantes, ce qui permet une étude plus adéquate de ces graphiques.

Il est important d'utiliser ces graphiques de façon complémentaire: le graphique double-masse donnant un aperçu de l'amplitude de la tendance possible et le graphique CUSUM donnant une meilleure appréciation des changements de pente. En effet, la hauteur du graphique des double-masse est défini par l'extrémité droite des cumuls et les écarts importants à cette droite seront un indice important quant à la possibilité d'une tendance significative. De son côté la hauteur du graphique des CUSUM est défini par l'amplitude. On aura donc toujours l'impression d'avoir une tendance significative, il faut donc se servir du CUSUM pour visualiser les changements de pente.

UN AUTRE TEST?

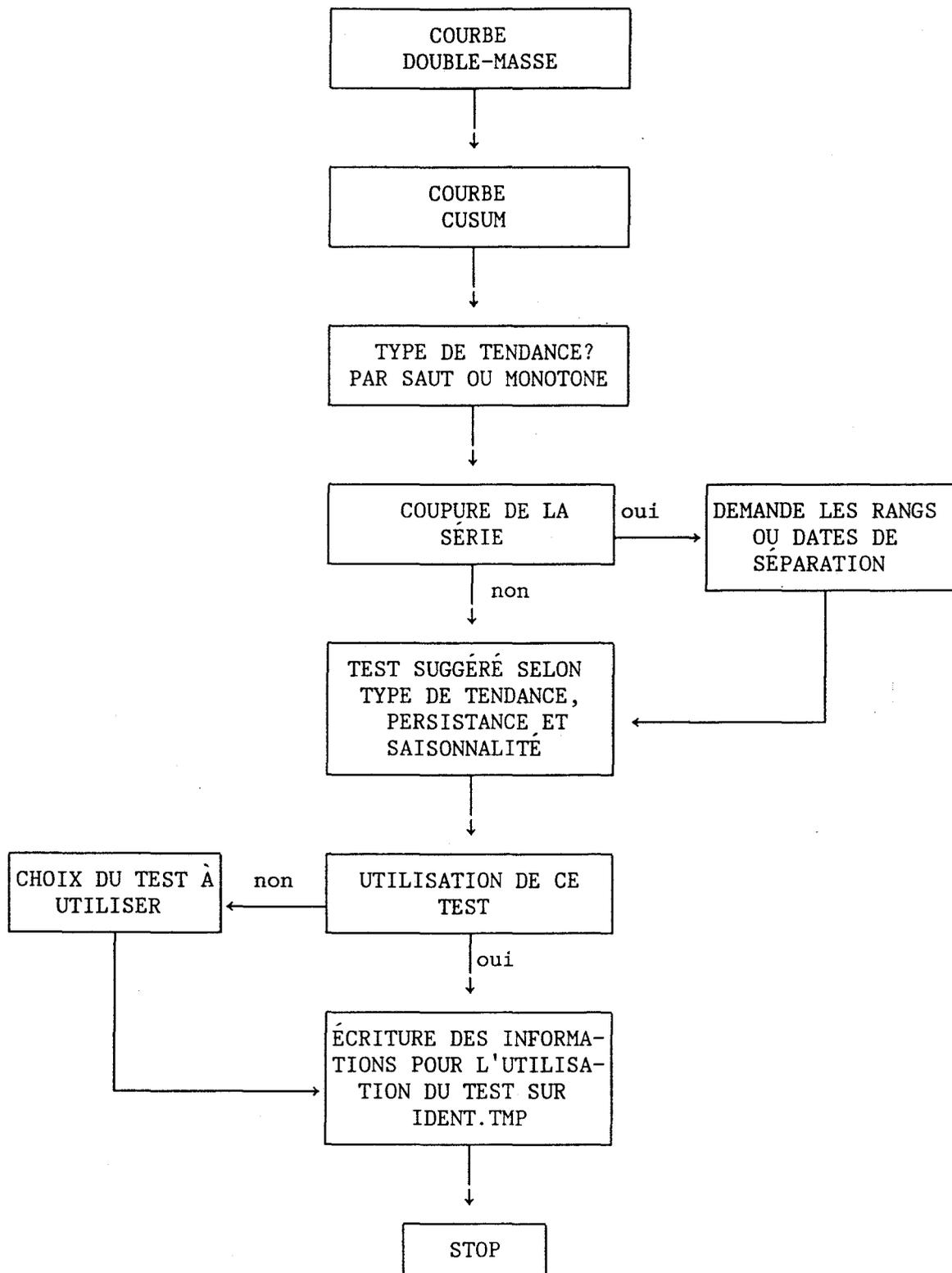
Bien que, de façon générale, le test proposé soit le test le plus approprié pour la série traitée, il peut cependant être intéressant dans certains cas de comparer le résultat avec d'autres tests. Lorsque l'on a détecté une persistance non-markovienne, les tests proposés utilisent l'adaptation de LETTENMAIER pour le cas d'une persistance markovienne. Il serait alors intéressant de voir les résultats des tests sans cette adaptation pour la persistance et de comparer les résultats quant à la détection ou non d'une tendance.

Pour ce qui est des saisonnalités, il peut arriver que les résultats de l'analyse de variance ne soient pas aussi convaincants que désirés soit à cause de la présence de valeurs aberrantes qui ne sont pas en accord avec l'hypothèse de normalité, soit à cause du rejet de l'égalité des variances. Dans de tels cas, il peut être intéressant de comparer les résultats des tests saisonniers et non-saisonniers.

La concordance des résultats de plusieurs tests permet de valider la conclusion devant les facteurs comme la persistance et la saisonnalité. Lorsque les résultats ne concordent pas, il faut alors porter un jugement le plus juste possible quant à la présence ou non de persistance et de saisonnalité afin d'utiliser le meilleur test possible.

Il peut aussi être opportun de choisir un autre test lorsque le test proposé est saisonnier et que le nombre d'observations est petit. En particulier pour les tests de MANN-WHITNEY saisonniers puisque leurs puissances ne sont pas connues et qu'il faut des données avant et après une certaine séparation.

9.1 Organigramme du programme P5_EXE



9.2 Contenu détaillé du programme P5.EXE

ACTION	RESTRICTIONS	SOUS-ROUTINES UTILISÉES
Courbe double-masse		DOUBMASS
Courbe CUSUM		CUSUM
Coupure de la série	Au choix de l'utilisateur.	
Suggestion d'un test	Déterminé à partir du type de tendance à traiter de la saisonnalité et de la persistance.	
Choix d'un test à utiliser	L'utilisateur n'est pas restreint à utiliser le test suggéré.	

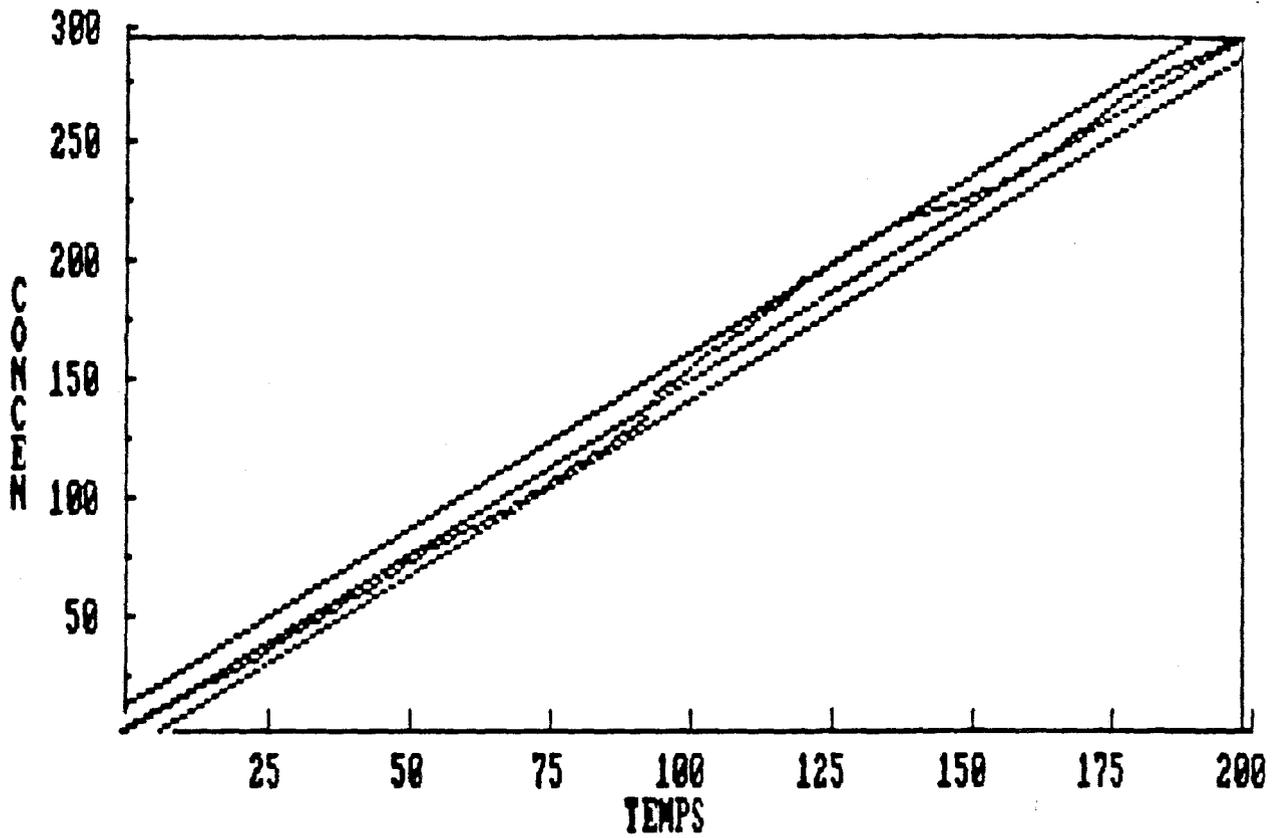
9.3 Figures du programme

Le programme contient 4 figures dont 2 seulement seront présentées selon le choix de l'utilisateur quant au traitement de la concentration ou de la charge.

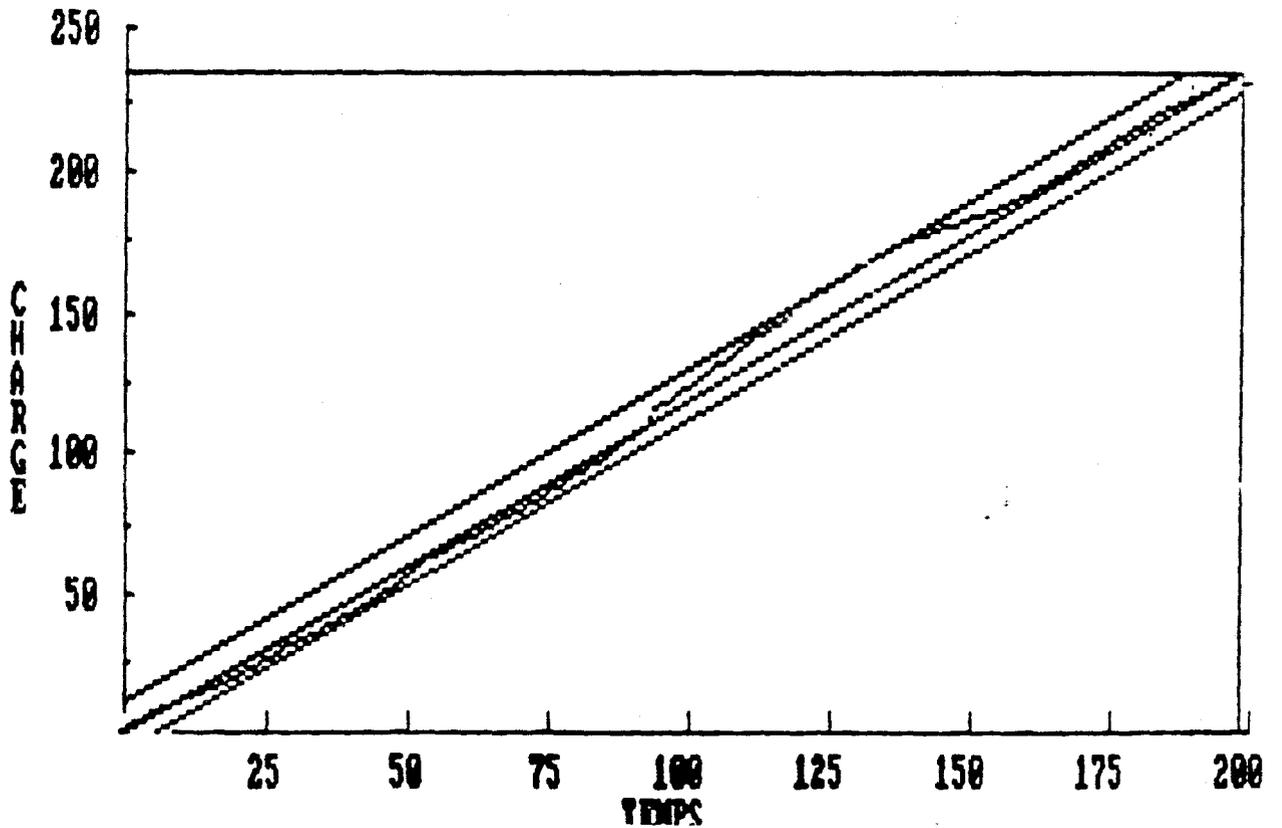
La figure 5.1 présente la courbe double-masse de la concentration du paramètre en fonction du temps. Cette courbe est construite à partir de la série équidistante et c'est ce qui la distingue de la figure 2.2, les structures de tendances seront cependant du même type. La figure 5.2 est présentée lorsque la charge est étudiée plutôt que la concentration.

La figure 5.3 présente la courbe CUSUM de la concentration du paramètre en fonction du temps. La série équidistante est également utilisée pour la construction de cette courbe. Ce graphique permet de détecter des points de ruptures pouvant servir de points de séparation dans le cas de l'utilisation d'un test de Mann-Whitney pour la détection d'une tendance par saut. La figure 5.4 est présentée lorsque la charge est étudiée plutôt que la concentration.

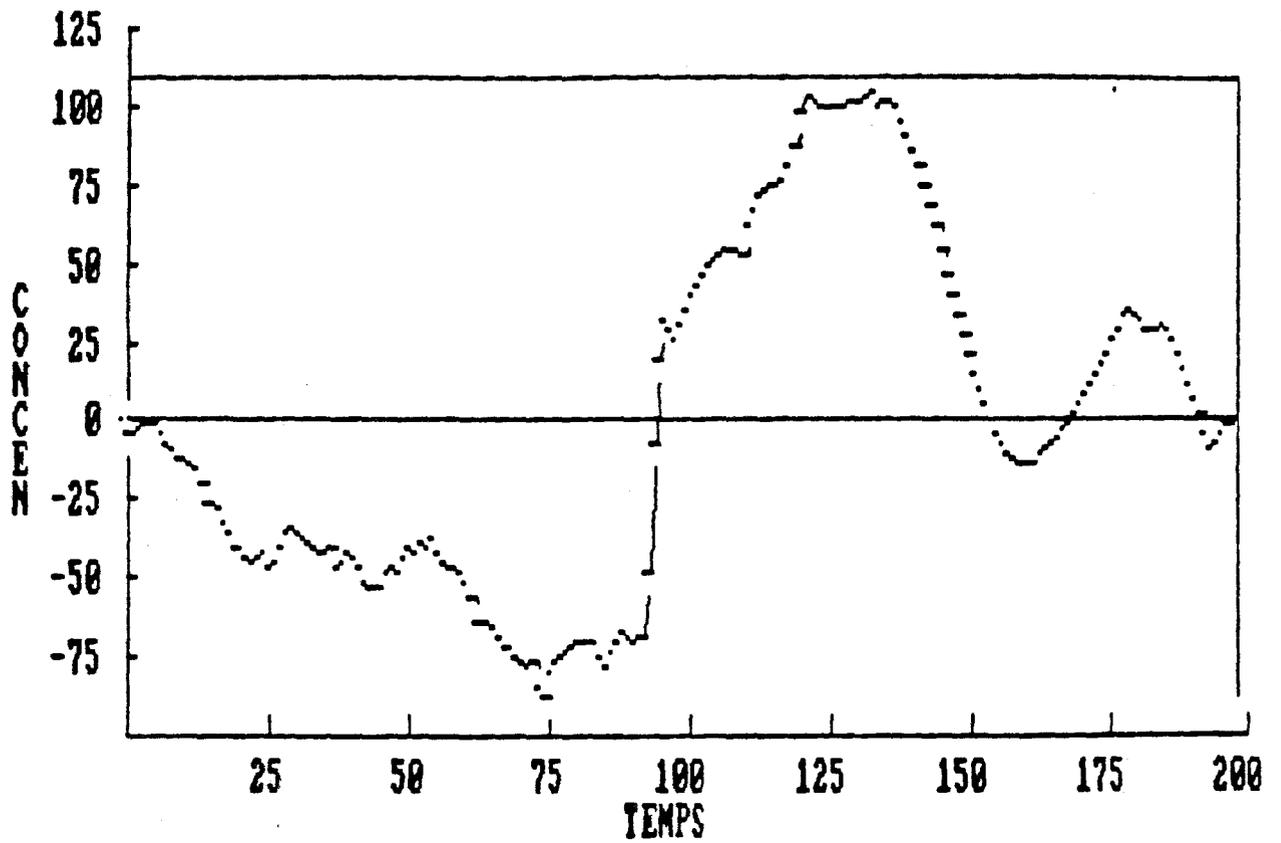
X 10E+02 FIGURE 5.1 : FONCTION DOUBLE-MASSE : C vs T, 9816CON



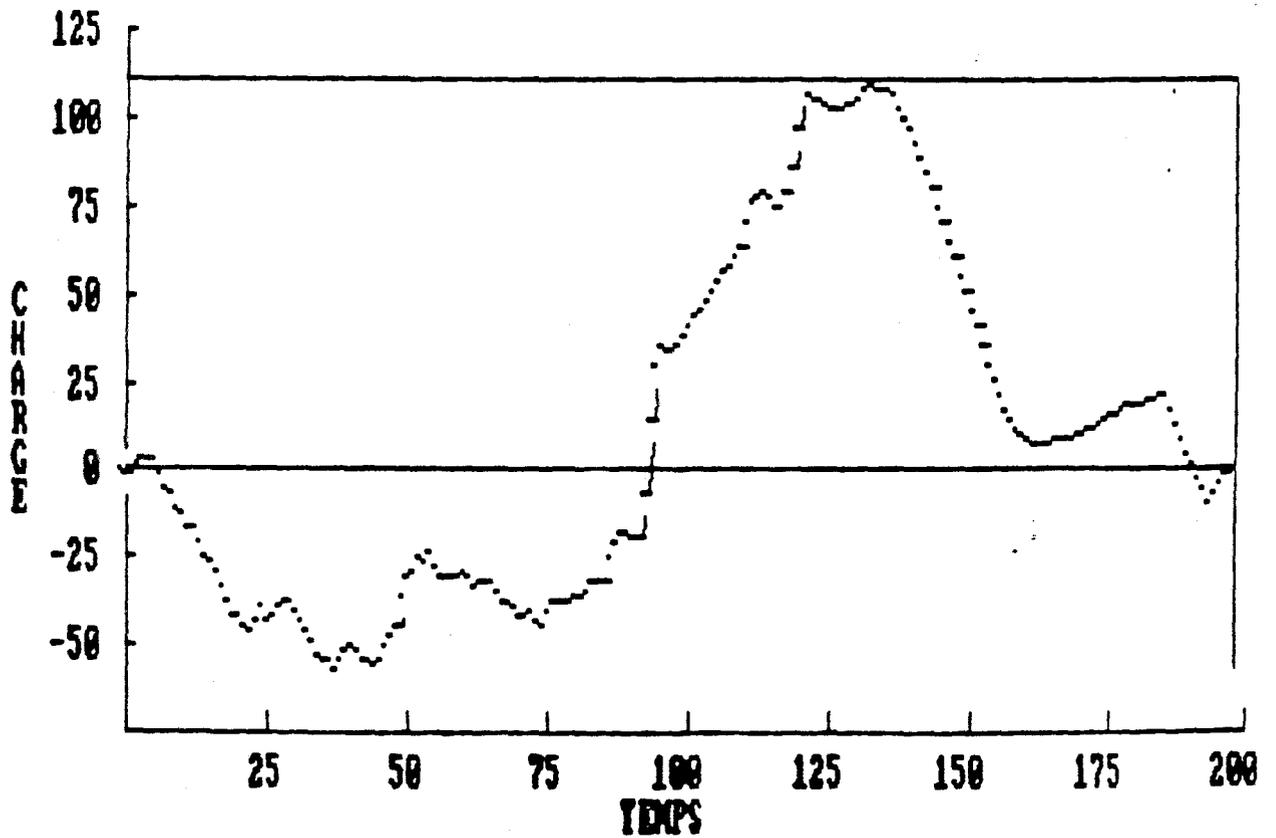
X 10E+11 FIGURE 5.2 : FONCTION DOUBLE-MASSE : L vs T, 9816CON



X 10E+01 FIGURE 5.3 : FONCTION CUSUM :C vs T, 9016CON



X 10E+10 FIGURE 5.4 : FONCTION CUSUM :L vs T, 9016CON



10. Tests

Douze tests sont disponibles à l'utilisateur. La majorité de ces tests sont des adaptations de tests non-paramétriques classiques afin de tenir compte des présences de saisonnalité et/ou de persistance. Six programmes permettent l'exécution des différents tests puisque des regroupements ont été faits de sorte que chaque programme exécute le test pour le cas sans persistance et le cas avec persistance correspondant.

Le tableau 10.1 présente le numéro et le nom de chaque test ainsi que le programme permettant l'exécution de ce test:

Tableau 10.1: Tests disponibles

Numéro du test	Nom du test	Nom du programme
1	MANN-WHITNEY	MW.EXE
2	MANN-WHITNEY/LETTENMAIER	MW.EXE
3	MANN-WHITNEY/SAISONNIER	MWS.EXE
4	MANN-WHITNEY SAIS/LETTENMAIER	MWS.EXE
5	KENDALL	KEN.EXE
6	SPEARMAN/LETTENMAIER	SP.EXE
7	KENDALL SAISONNIER	KENS.EXE
8	HIRSCH ET SLACK	KENS.EXE
9	FOSTER ET STUART 1	FS.EXE
10	FOSTER ET STUART 2	FS.EXE
11	TAU DE KENDALL	KEN.EXE
12	SPEARMAN	SP.EXE

Tous les renseignements sur la construction et l'application de ces tests sont présentés dans la partie 14. Des fiches techniques sont alors présentées pour chacun.

11. Synthèse (P6.EXE)

Ce programme effectue la synthèse de tous les résultats obtenus sur la série et, présente sous une forme paramétrique, les résultats obtenus, avec les dates de changement, les niveaux des paramètres et les pentes des tendances.

GÉNÉRALITÉS

CRITÈRE D'AJUSTEMENT

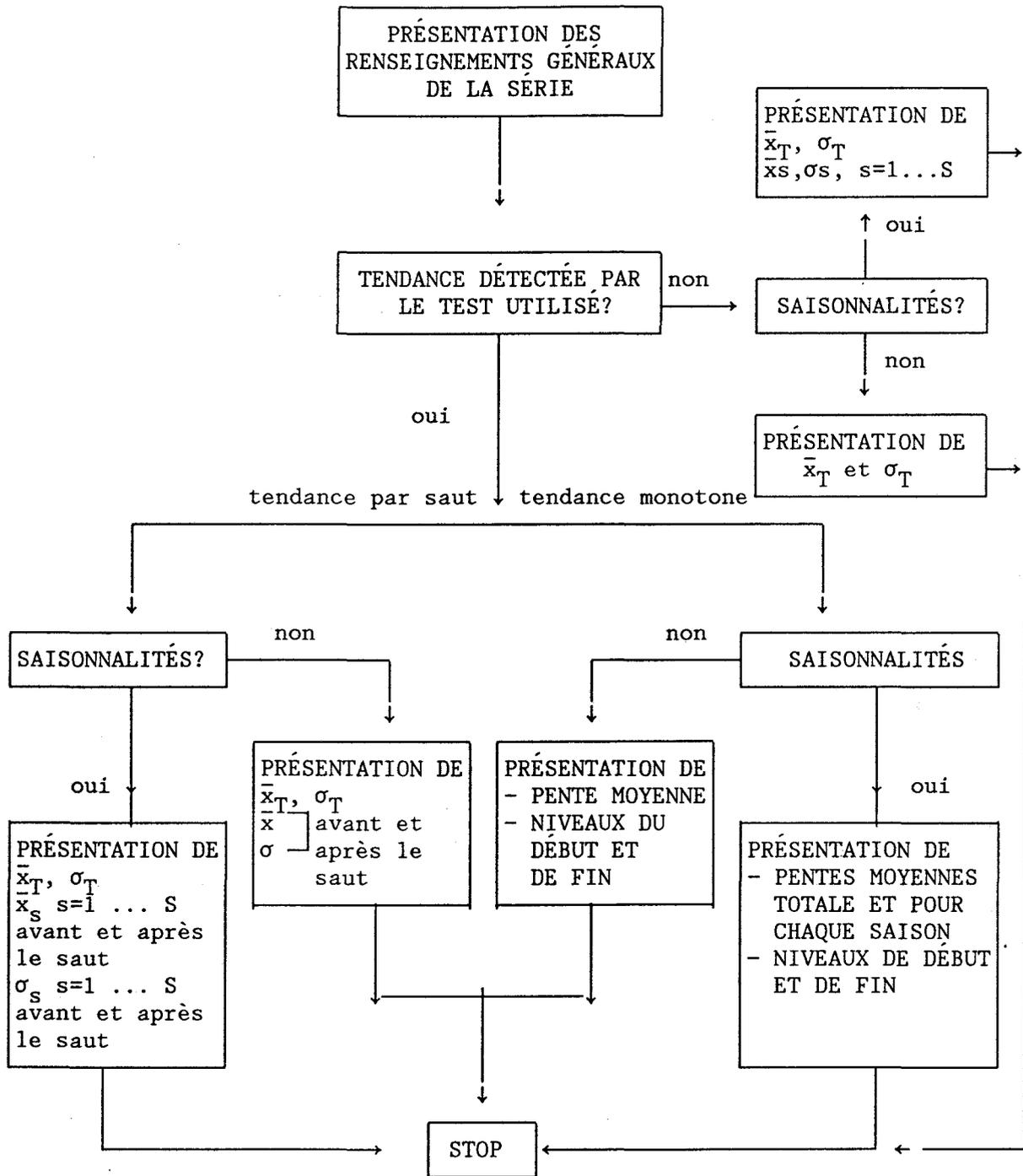
Le critère d'ajustement et l'étude graphique utilisés à la fin de P6 permettent de juger la qualité de l'utilisation d'un modèle de tendance par saut ou monotone. Le critère d'ajustement est la racine carrée de l'erreur quadratique moyenne par rapport au modèle choisi. On prend donc les écarts à la droite de régression pour un modèle de tendance monotone et les écarts aux moyennes avant et après la séparation dans le cas d'un modèle par saut. Plus l'erreur quadratique moyenne (EQM) est petite, meilleur est l'ajustement. L'élimination des valeurs extrêmes dans P2 réduira généralement cette EQM, mais ne devrait pas être utilisée dans ce but.

PARAMÉTRIQUE VS NON-PARAMÉTRIQUE

Étant donné que les tests de détection de tendance utilisés dans le logiciel sont tous non-paramétriques et que les études effectuées dans P6 sont paramétriques, il peut parfois y avoir des discordances dans les résultats. Par exemple, on peut détecter une tendance monotone croissante à l'aide du test de KENDALL (rang du paramètre augmente avec le temps), alors que la régression nous donne une pente négative (valeur du paramètre diminue avec le temps). Dans de tels cas, la présence d'une ou de plusieurs valeurs aberrantes est généralement la cause de tels résultats contradictoires. L'utilisateur peut alors reprendre l'exécution à la partie P2 pour éliminer la ou les valeurs aberrantes s'il désire des résultats concordants.

Un tel phénomène est possible à cause de la faible influence qu'ont les valeurs aberrantes sur les méthodes non-paramétriques comparativement aux méthodes paramétriques. Il est donc toujours préférable d'accorder plus d'importance aux résultats de la méthode non-paramétrique.

11.1 Organigramme du programme P6.EXE



\bar{x}_T = moyenne de la série totale
 σ_T = écart-type de la série totale
 \bar{x} = moyenne
 σ = écart-type
 S = nombre de saisons

11.2 Contenu détaillé du programme P6.EXE

ACTION	RESTRICTION	SOUS-ROUTINES UTILISÉES
Présentation des renseignements généraux de la série		
Présentation des statistiques globales	Dans le cas où il n'y a pas de tendance détectée et qu'il n'y a pas de saisonnalité	STATGLOB
Présentation des statistiques globales	Dans le cas où il n'y a pas de tendance détectée et qu'il y a saisonnalité	STATSAIS
Présentation des statistiques avec tendance par saut	Si on a choisi le test 1 ou 2 (Mann-Whitney) et qu'il y a tendance détectée	STAMW
Présentation des statistiques avec tendance par saut et saisonnalité	Si on a choisi le test 3 ou 4 (Mann-Whitney) et qu'il y a tendance détectée	STAMWSAIS
Présentation de la pente moyenne et des niveaux de début et de fin	S'il y a tendance détectée et qu'il n'y a pas de saisonnalité	MONOTONE
Présentation des pentes moyennes pour chaque saison, et des niveaux de début et de fin	S'il y a tendance détectée et qu'il y a saisonnalité	MONOSAIS

11.3 Sortie des résultats du programme

Le tableau 11.1 présente une sortie typique du programme P6. Plusieurs sorties distinctes sont possibles selon la présence ou non de saisonnalités et selon le type de tendance étudié. Toutes les sorties en mode texte sont faites simultanément à l'écran et dans un fichier (SYNTHESE.P6). Le fichier SYNTHESE.P6 doit exister à l'exécution de P6 et chaque passage dans cette partie du logiciel ajoutera des résultats à ce même fichier. Il est donc important de bien traiter ce fichier. Voici quelques recommandations à ce propos:

- Si l'utilisateur interrompt le traitement avant la fin de P6, il est possible que le fichier SYNTHESE.P6 n'existe plus. Il est cependant possible de le retrouver puisque TAMPON.P6 sera une copie exacte du fichier SYNTHESE.P6 avant l'entrée dans P6.
- Un fichier vide appelé BIDON.P6 devrait toujours exister lors d'utilisation à l'aide des programmes "batch" car lorsqu'aucun fichier SYNTHESE.P6 n'est présent, le logiciel en créera un à partir de BIDON.P6.
- Pour une édition de SYNTHESE.P6 à l'aide de SIDEKICK il faut que ce fichier ait moins de 45K. Il faudra donc renommer SYNTHESE.P6 avec des noms plus précis aussi souvent que possible.

La seule différence entre ce qui est vu à l'écran et ce qui est écrit dans SYNTHESE.P6 est la présentation à l'écran du graphique illustrant la qualité de l'ajustement. La figure 11.1 présente un graphique pour un ajustement par saut alors que la figure 11.2 présente un graphique pour un ajustement à l'aide d'une tendance monotone. Dans les cas où il n'y a pas de tendance seule la moyenne générale est représentée à l'aide d'une droite.

Tableau 11.1: Sortie de P6

S Y N T H E S E SESSION DU 9 JUILLET 1987
 NOM DU FICHIER : MC0138.ALC SERIE TRAITEE : CONCENTRATIONS
 SERIE ORIGINALE : DATE DE LA PREMIERE OBSERVATION : 15/ 8/72
DATE DE LA DERNIERE OBSERVATION : 5/ 6/78
NOMBRE D'OBSERVATIONS : 74
NOMBRE DE MINIMA RETIRES : 0
NOMBRE DE MAXIMA RETIRES : 0
 ANNEES A FAIBLE FREQUENCE : AUCUNE
 METHODE EMPLOYEE POUR COMPLETER LA SERIE : MOYENNES SAISONNIERES
 INTERVALLE DE TRAVAIL : MOIS DE DEPART : 1
INTERVALLE (EN MOIS) : 1
NOMBRE D'INTERVALLES PAR ANNEE : 12
 PERSISTANCE : MARKOVIENNE SAISONNALITE : OUI UTILISEE : OUI
 SERIE DE TRAVAIL : DATE DE DEPART : 15/ 8/72 (RANG = 1)
DATE DE FIN : 15/ 6/78 (RANG = 71)
NOMBRE DE POINTS : 71
 TEST SUGGERE : HIRSCH ET SLACK
 TEST UTILISE : HIRSCH ET SLACK
 RESULTATS DU TEST UTILISE : VALEUR DE LA STATISTIQUE : 6.0000
VALEUR DU TEST : .1914
SIGNIFICATION DU TEST : .5759
COMMENTAIRE : PAS DE TENDANCE DETECTEE
 MOYENNE GLOBALE : 91.931
 ECART-TYPE : 7.788
 ECART-TYPE DE LA MOYENNE : .588

SAISON	MOYENNE	ECART-TYPE	EC-TYPE DE LA MOYENNE
MOIS 1	95.100	1.753	.455
MOIS 2	95.550	4.720	1.226
MOIS 3	94.297	5.100	1.324
MOIS 4	96.836	4.816	1.251
MOIS 5	97.500	6.592	1.712
MOIS 6	93.960	4.188	1.088
MOIS 7	93.075	3.295	.937
MOIS 8	80.025	3.023	.785
MOIS 9	84.175	6.954	1.806
MOIS 10	81.963	5.667	1.472
MOIS 11	94.080	5.754	1.494
MOIS 12	96.800	5.634	1.463

 ADEQUATION DU MODELE DE TENDANCE MONOTONE AUX DONNEES
 POUR L'INTERVALLE DE TRAVAIL CHOISI :
RMSE : 7.788 mg/l

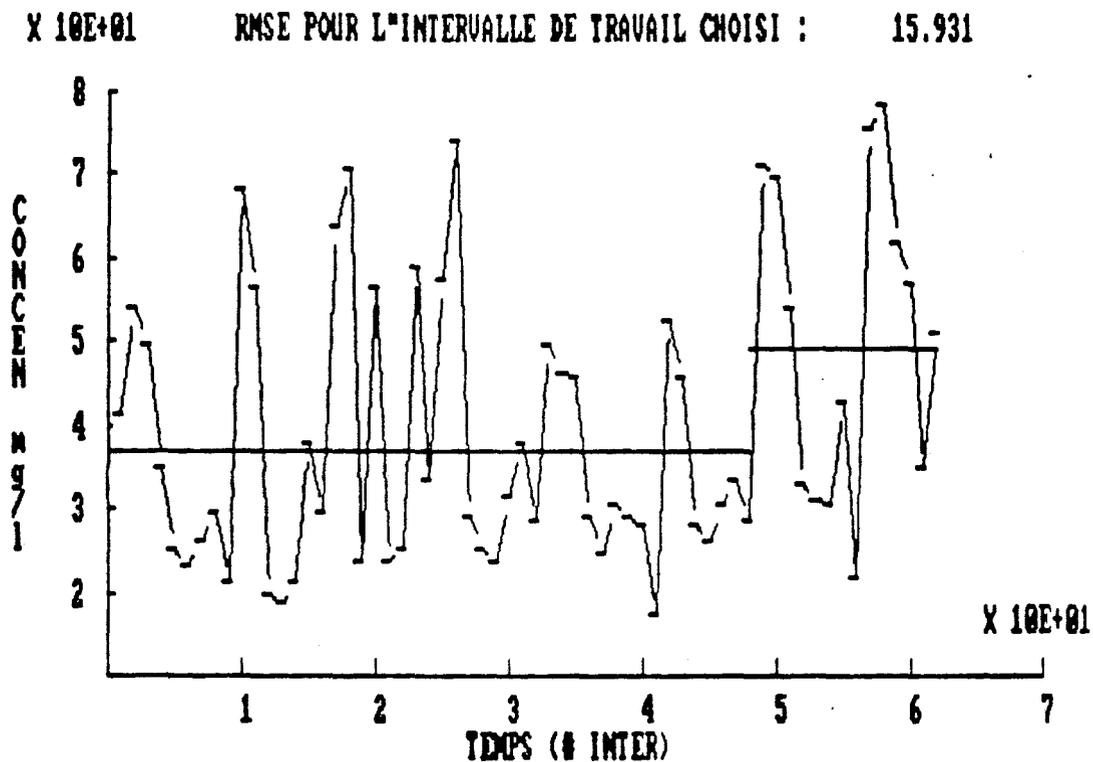


Figure 11.1: Ajustement à l'aide d'une tendance par saut.

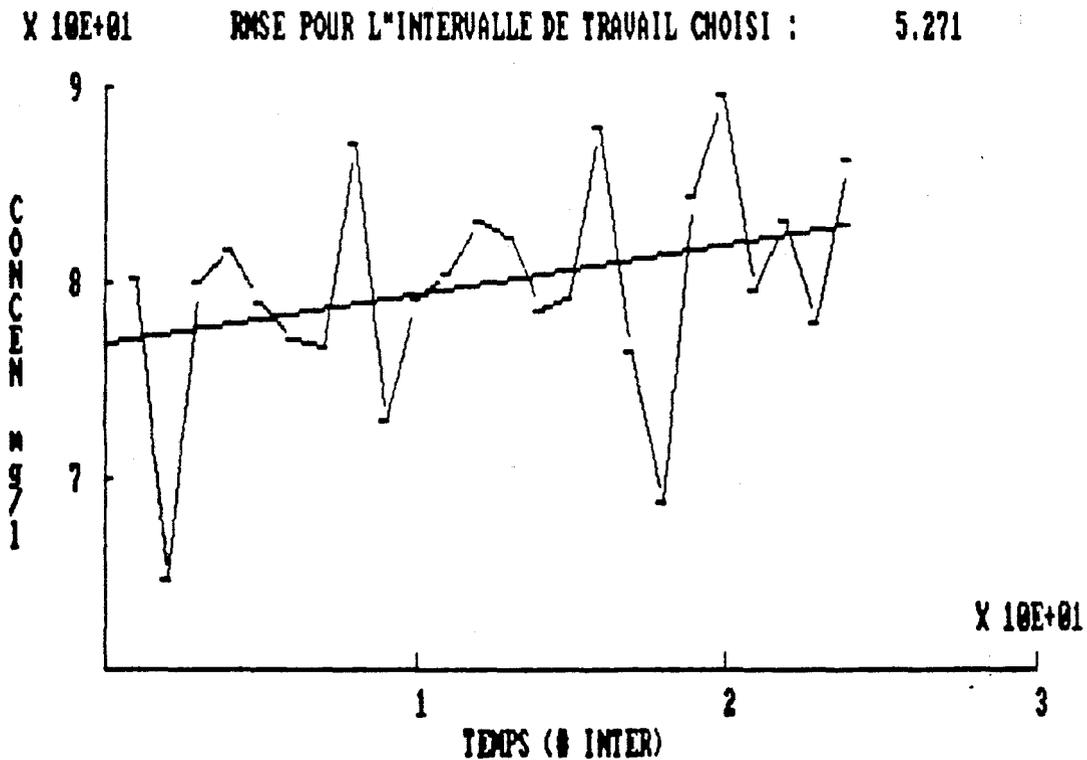


Figure 11.2: Ajustement à l'aide d'une tendance monotone.

12. ÉTUDE DES DÉPASSEMENTS (P7.EXE)

Le but de ce programme est d'étudier la fréquence, l'intensité et la durée des valeurs élevées des séries mesurées (concentrations ou charges) définies par rapport à une norme environnementale ou réglementaire (Figure 1). Il permet aussi de comparer deux sous-populations de la série étudiée. La définition de ces deux sous-ensembles peut être le résultat de l'analyse du logiciel complet (DETECT) exécuté antérieurement (avant et après l'intervention détectée, par exemple), ou bien complètement arbitraire (données récentes vs données historiques, mois d'été vs mois d'hiver, données de crues vs données d'étiage). Dans tous les cas, seule l'exécution du programme de préparation P1 est nécessaire.

12.1 Définition des deux sous populations

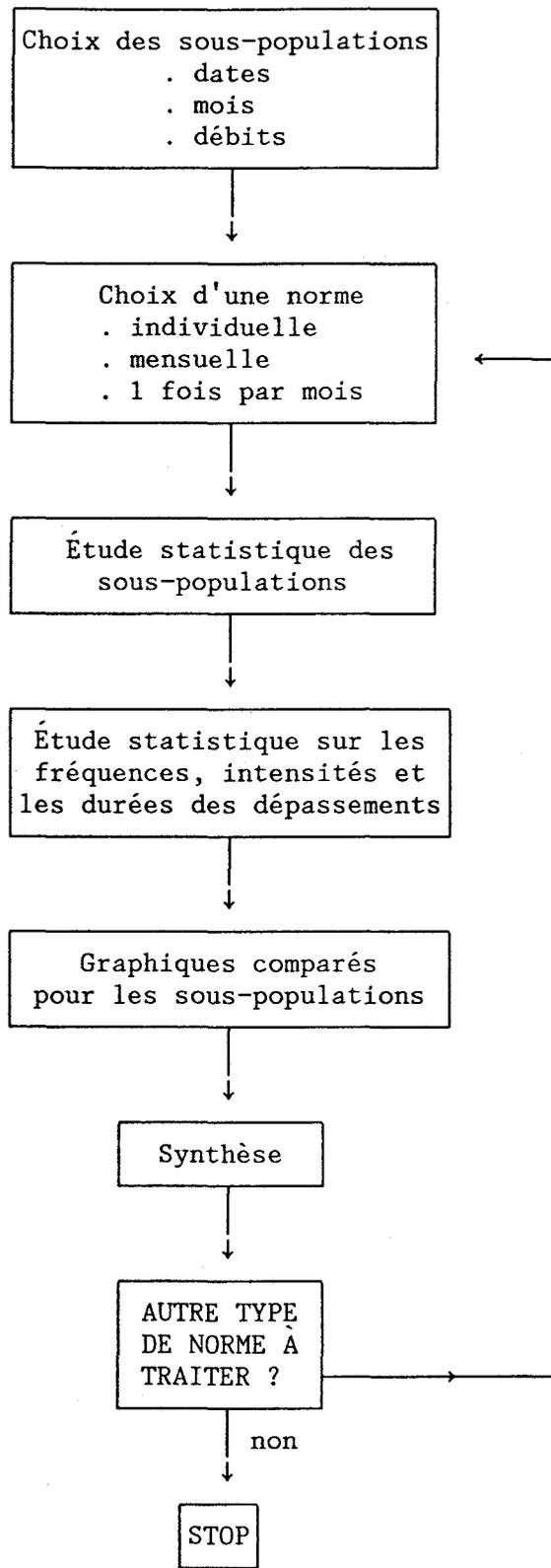
Les sous-populations sont définies par intersections booléennes selon les dates (début-fin), selon des mois définis et selon des classes de débits.

12.2 Choix d'une norme ou d'un seuil

Une fois les sous-populations définies, l'utilisateur doit choisir une norme de référence. Trois normes sont proposées:

- a) une norme générale ou individuelle à ne jamais dépasser. Elle s'applique à chaque valeur des populations retenues;
- b) une norme mensuelle qui s'applique aux moyennes mensuelles des valeurs retenues;
- c) une norme individuelle que l'on peut dépasser une fois par mois, elle s'applique à chaque valeur des populations retenues, la valeur la plus élevée de chaque mois étant éliminée comme dépassement.

Figure 12: Organigramme du programme P7.EXE



12.3 Traitement

Une fois ce cadre d'analyse établi, le programme calcule et imprime le bilan statistique de chaque sous-population (individuelle ou mensuelle);

- nombre de valeurs présentes
- moyenne
- écart-type
- médiane
- nombre de dépassements
- pourcentage de dépassements

Un tableau présente ensuite les statistiques concernant les durées des dépassements dans chaque sous-population:

- longueur du dépassement et fréquence associée
- durée moyenne des dépassements
- écart-type associé

On présente alors une série de graphiques concernant les distributions et distributions cumulées comparées des deux sous-populations choisies. Puis viennent les graphiques comparant les dépassements reliés à chaque sous-population. Sur ces derniers, la valeur de la norme choisie met en évidence l'importance relative des valeurs hors-norme. Finalement, un dernier graphique est présenté pour comparer les durées des dépassements successifs.

Tout comme le programme P6, le programme P7 écrit dans un fichier l'ensemble des résultats pertinents à l'analyse. Ce fichier est appelé SYNTHESE.P7 et chaque traitement à l'aide de P7 amènera une écriture complémentaire de résultats dans ce fichier. Les tableaux 1 à 3 présentent une sortie typique des résultats obtenus par P7 y compris le fichier de synthèse.

De plus, lors du traitement à l'aide de P7, un ensemble de graphiques (Figures 7.1 à 7.6) est présenté à l'utilisateur à l'écran. Ils seront également imprimés si l'utilisateur a choisi une telle option dans P1.

Tableau 1: Déroulement du programme P7.

NOM DU FICHIER: `entree.dat` SERIE TRAITEE: CONCENTRATIONS # POINTS: 109
ENTRE LE 1/ 6/69 ET LE 12/10/82 DEBITS NON-DISPONIBLES

VOULEZ-VOUS CREER LES SOUS-POPULATIONS SELON LES DATES ? (O/N) : o

ENTREZ LES DATES DELIMITANT LA PREMIERE SOUS-POPULATION A COMPARER :

DEBUT --> MOIS : 6
ANNEE : 1969
FIN ----> MOIS : 8
ANNEE : 1975

ENTREZ LES DATES DELIMITANT LA SECONDE SOUS-POPULATION A COMPARER :

DEBUT --> MOIS : 9
ANNEE : 1975
FIN ----> MOIS : 10
ANNEE : 1982

NOM DU FICHIER: `entree.dat` SERIE TRAITEE: CONCENTRATIONS # POINTS: 109
ENTRE LE 1/ 6/69 ET LE 12/10/82 DEBITS NON-DISPONIBLES

VOULEZ-VOUS SPECIFIER LES SOUS-POPULATIONS PAR RAPPORT AUX MOIS ? (O/N) : n
QUELLE NORME DESIREZ-VOUS TRAITER ?

- 1) NORME INDIVIDUELLE
- 2) NORME MENSUELLE
- 3) 1 FOIS PAR MOIS
- 4) FIN

VOTRE CHOIX : 1

Tableau 2: Déroulement du programme P7 (suite).

EXTREMA DES SOUS-POPULATIONS : 3.00 ET 12.70			
ENTREZ LA NORME INDIVIDUELLE A NE PAS DEPASSER : 9			
	SOUS-POP 1	SOUS-POP 2	
NOMBRE DE VALEURS :	60	49	
MOYENNE :	9.560	8.549	
ECART-TYPE :	1.823	1.305	
MEDIANE :	9.500	8.400	
NOMBRE DE DEPASSEMENTS :	35	13	
% DE DEPASSEMENTS :	58.333	26.531	

DEPASSEMENTS

LONGUEUR DU DEPASSEMENT	FREQUENCE SOUS-POP1	FREQUENCE SOUS-POP2
0	25	36
1	6	1
2	1	1
3	0	2
4	0	1
5	0	0
6	0	0
7	1	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	1	0

DUREE MOYENNE D'UN DEPASS. :	3.89	2.60
ECART-TYPE ASSOCIE :	5.99	1.02

Tableau 3: Déroulement du programme P7 (suite).

S Y N T H E S E D E P 7		SESSION DU 19 FEVRIER 1988	
NOM DU FICHIER :	entree.dat	SERIE TRAITEE :	CONCENTRATIONS
SERIE TRAITEE :	DATE DE LA PREMIERE OBSERVATION :	1/ 6/69	
	DATE DE LA DERNIERE OBSERVATION :	12/10/82	
	NOMBRE D'OBSERVATIONS :	109	
NORME UTILISEE :	(INDIVIDUELLE)	:	9.000 mg/L

	SOUS-POP 1		SOUS-POP 2
PERIODE :	6/1969 - 8/1975	:	9/1975 - 10/1982
MOIS UTILISES :	TOUS	:	TOUS
DEBIT CONSIDERES :	TOUS	:	TOUS
NOMBRE DE VALEURS :	60	:	49
MOYENNE DES DONNEES :	9.56 mg/L (1.82)	:	8.55 mg/L (1.30)
NOMBRE DE DEPASSEMENTS :	35 (58.33%)	:	13 (26.53%)
DEPASSEMENT MOYEN :	10.74 mg/L (1.03)	:	10.22 mg/L (.75)
DUREE MOYENNE :	3.89 int (5.99)	:	2.60 int (1.02)

DISTRIBUTION DES SOUS-POPULATIONS :

SOUS-POP 1 :	3.0	1.7%	4.9	3.3%	6.9	28.3%	8.8	38.3%	10.8	28.3%	12.7	mg/L
SOUS-POP 2 :		.0%		8.2%		55.1%		28.6%		8.2%		

REPARTITION DES DEPASSEMENTS SELON LA NORME :

SOUS-POP 1 :	41.7%	58.3%	.0%	.0%	.0%
		N	2N	3N	4N ET PLUS
SOUS-POP 2 :	73.5%	26.5%	.0%	.0%	.0%

REPARTITION DES DUREES DE DEPASSEMENT :

SOUS-POP 1 :	0	41.7	1	10.0	2	3.3	3	.0	4	45.0	ET PLUS
SOUS-POP 2 :		73.5		2.0		4.1		12.2		8.2	

QUELLE NORME DESIREZ-VOUS TRAITER ?

- 1) NORME INDIVIDUELLE
- 2) NORME MENSUELLE
- 3) 1 FOIS PAR MOIS
- 4) FIN

VOTRE CHOIX : 4

VOULEZ-VOUS VOIR LES GRAPHIQUES (O/N)? 0
X 10E+01 **FIGURE 7.1** DISTRIBUTION DES SOUS-POP , entree.dat
SOUS-POP2 MOY: 9.56 ECART-TYPE: 1.82

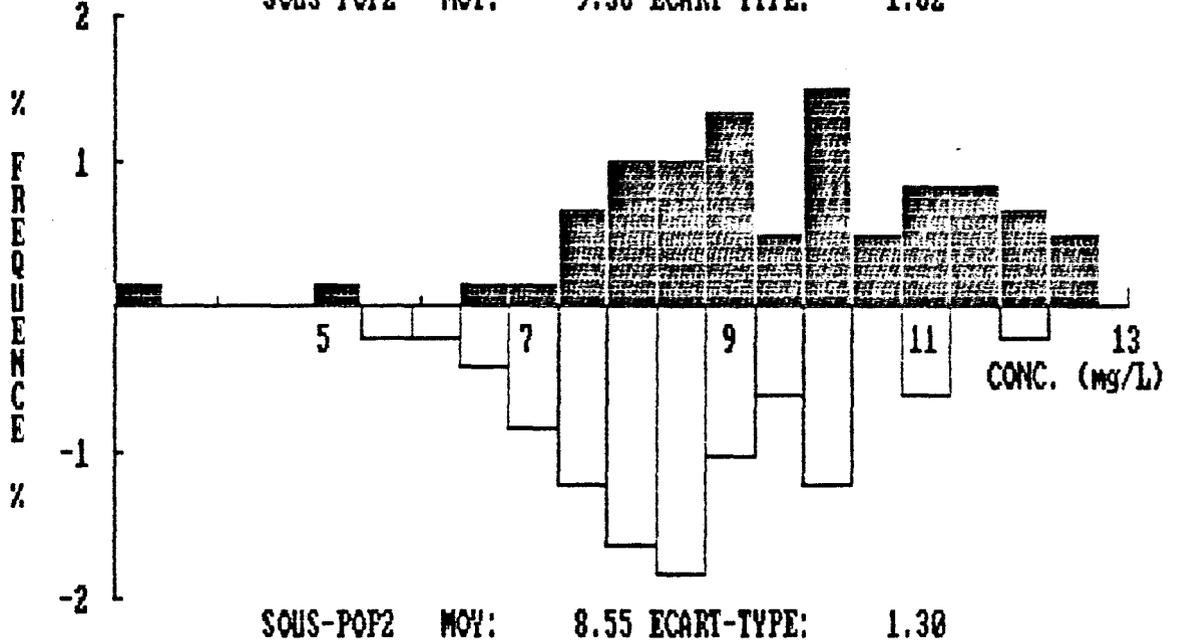


FIGURE 7.2 DISTRIBUTION CUMULEE DES SOUS-POP , entree.dat

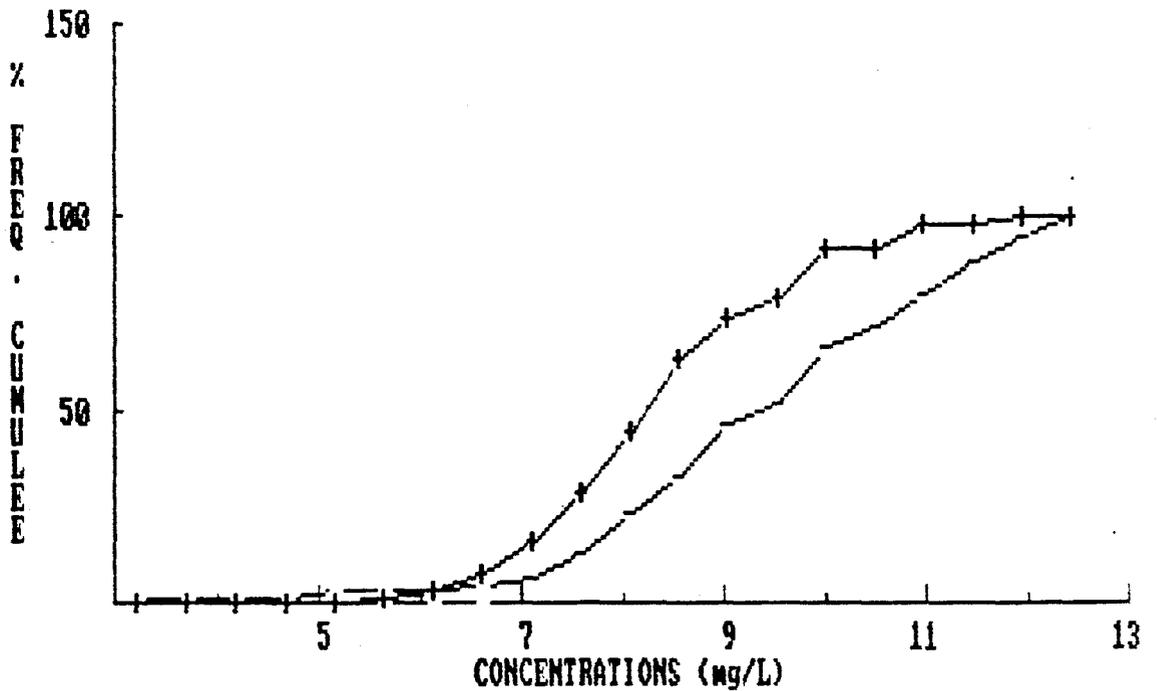
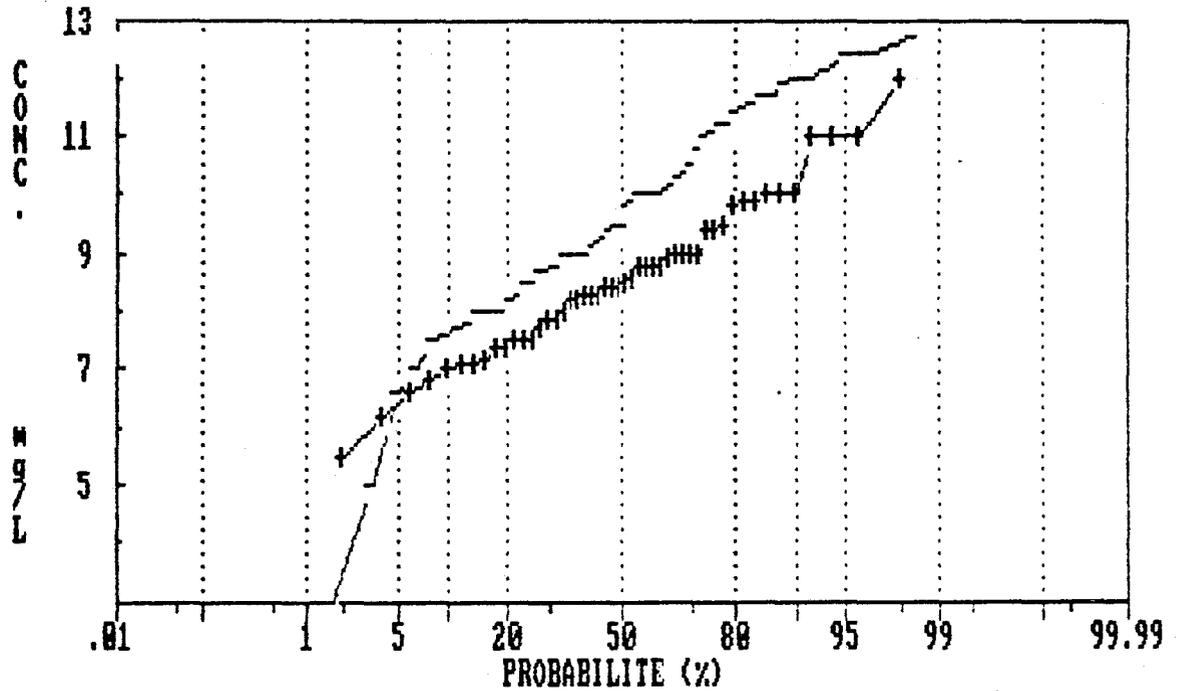


FIGURE 7.3 DISTRIBUTION CUMULEE DES SOUS-POP , entree.dat



X 10E+01 FIGURE 7.4 DISTRIBUTION DES DEPASSEMENTS , entree.dat
SOUS-POP2 MOY: 10.74 ECART-TYPE: 1.03

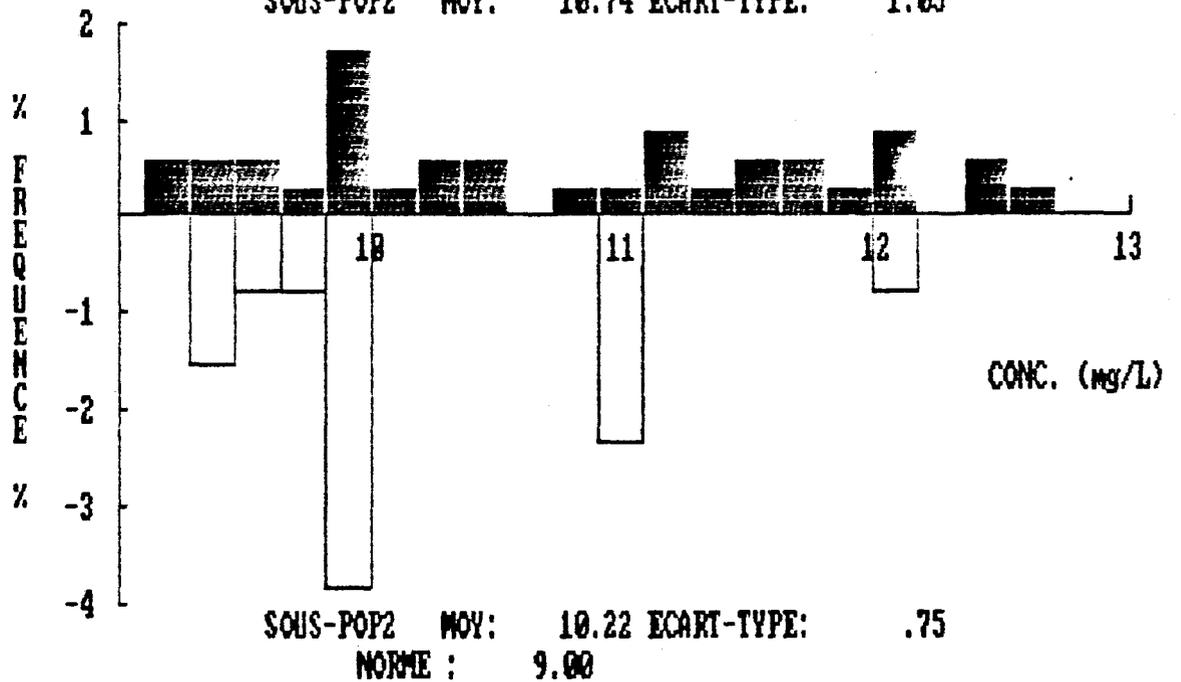
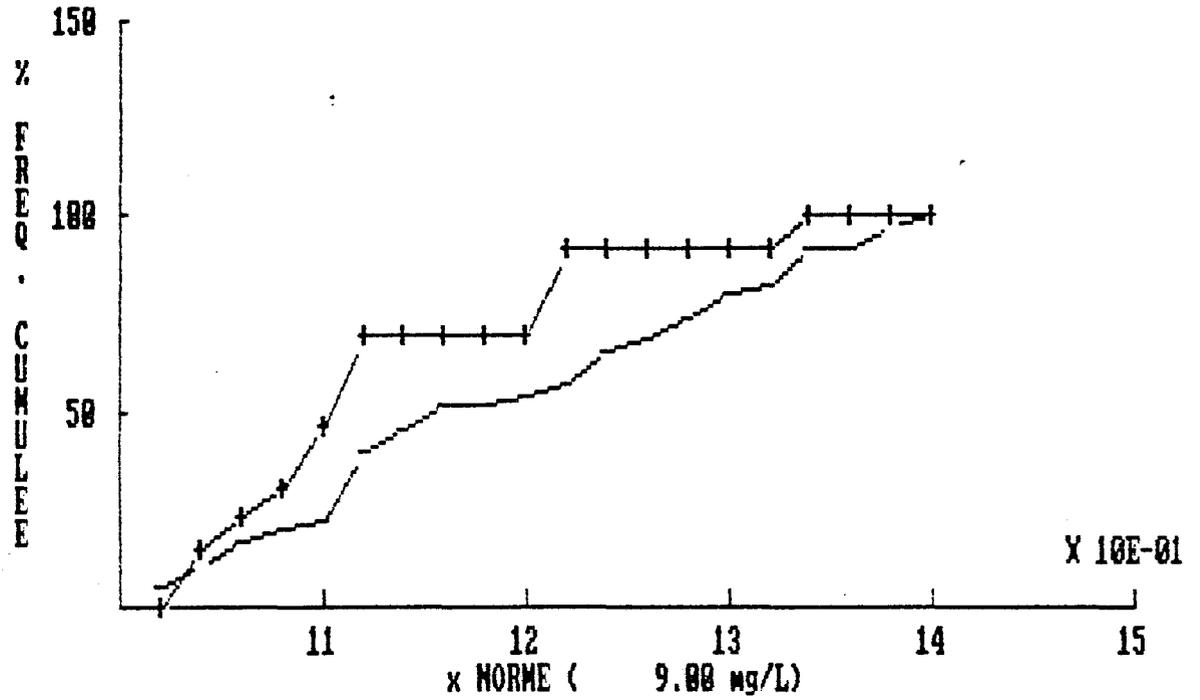


FIGURE 7.5 DISTR. CUMULEE DES DEPASSEMENTS , entree.dat



X 10E-01 FIGURE 7.6 DISTR. CUMULEE DES DEPASSEMENTS , entree.dat

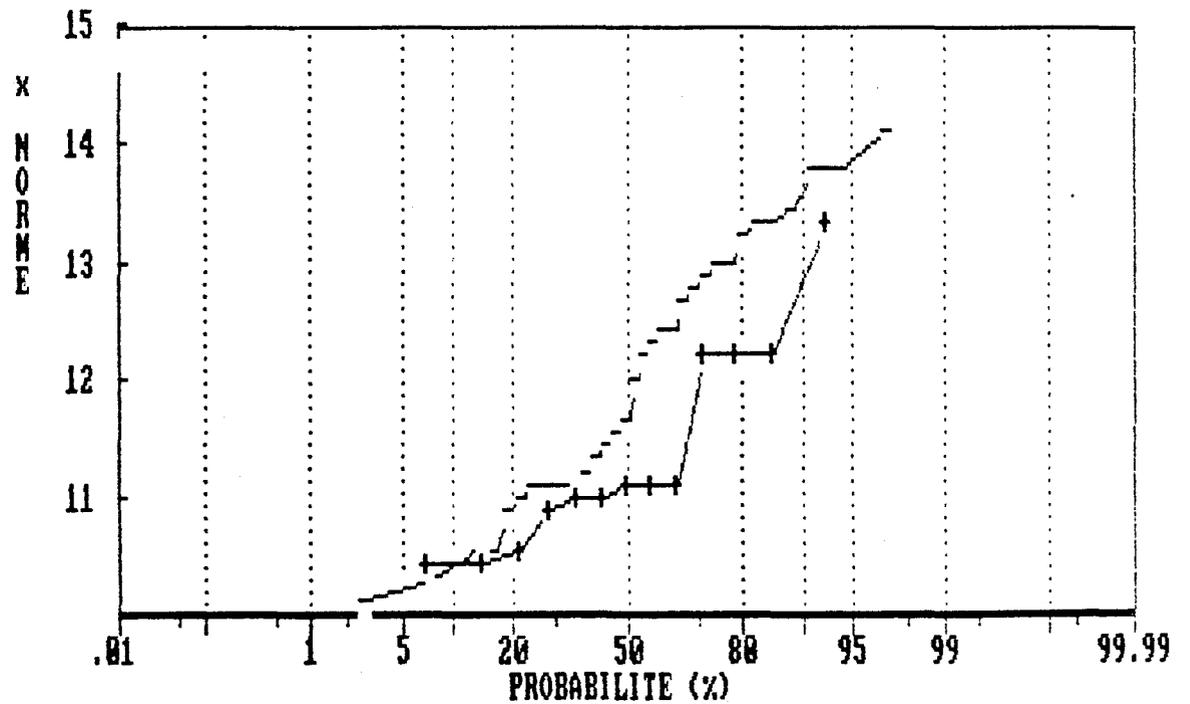
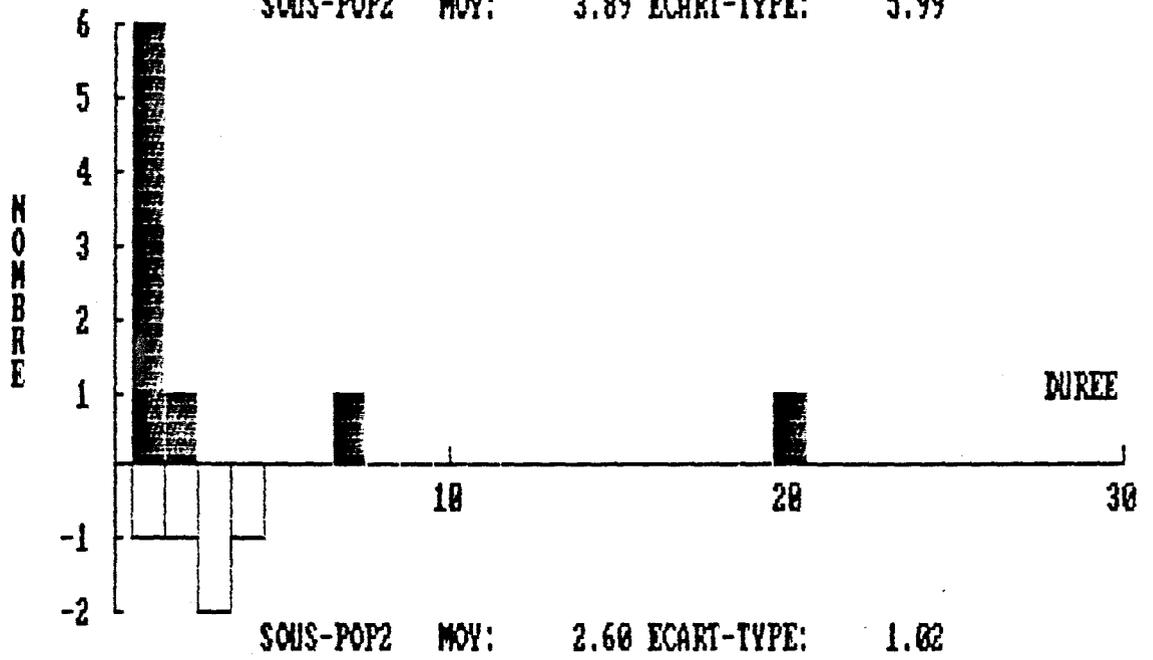
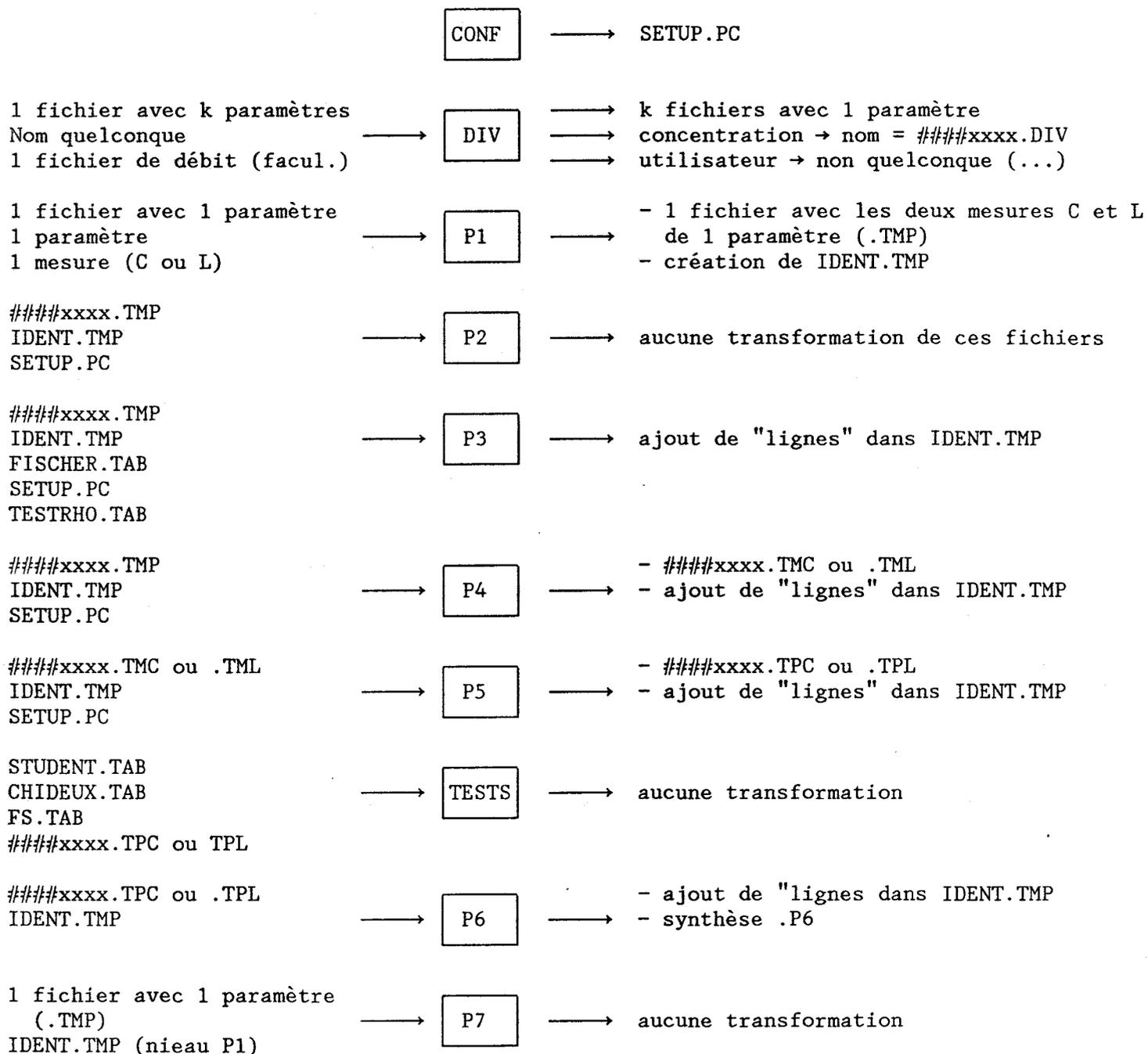


FIGURE 7.7 DUREES DES DEPASSEMENTS, entree.dat
SOUS-POP2 MOY: 3.89 ECART-TYPE: 5.99



13. Structure générale des fichiers

13.1 Organigramme général de l'utilisation des fichiers



13.2 Structure des fichiers

1 SETUP.PC

origine: CONF.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT
1	ITMODE, ITFORE, ITBACK, ITCOLOR	Variables pour la configuration en mode texte	4I2
2	IGMODE, IGBACK, IPAL, IGBORD, IGCOLOR, ILINE, ILETR, ISEC, KIRSYM, ISYMBL, JROWI, JROW2, JCOL1, JCOL2	Variables pour la configuration en mode graphique	9I2, 5I3

2 "FICHIER".TMP

origine: P1.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT
i 1 ≤ i ≤ NOBS	AN, MO, JR CONC, DEB, CHARGE RANG	Date de l'observation, concentration, débit et charge Rang relatif à partir de la date de la lère observation	I2, 1X, I2, 1X, I2, 1X F12.6, IX, F12.6, IX, F20.5, 17X, I5

3 "FICHIER".TMC/TML

origine: P4.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT
i 1 ≤ i ≤ LS	.TMC : CONC .TML : CHARGE	Concentration ou charge; (série une fois les trous bouchés, sans les valeurs absentes au début et à la fin)	.TMC : F10.3 .TML : F20.3

4 "FICHIER" .TPC/TPL

origine: P5.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT
i 1 ≤ i ≤ LONG	.TPC : CONC .TPL : CHARGE	Concentration ou charge (série ou sous-série extraite du fichier .TMC/.TML)	.TPC : F10.3 .TPL : F20.3

5

IDENT.TMP

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT	ORIGINE
1	NOMFICH LG	Nom du fichier d'entrée Longueur de NOMFICH (en nombre de lettres avant l'extension).	A12,1X,I1	P1.FOR
2	FICHDEB IDEB	Nom du fichier de débits journaliers(si utilisé). En cas d'absence, FICHDEB = caractères blancs =1 si les débits sont dans le même fichier d'entrée =2 si on doit faire référence au fichier de débits journaliers =3 si les débits ne sont pas utilisés	A12,1X,I1	P1.FOR
3	ITYP IMPRIM	=0 si on utilise les concentrations =1 si on préfère les charges =0 si on ne fait pas d'impressions =1 si l'utilisateur veut imprimer chacun des graphiques	I1,1X,I1	P1.FOR
4	LOW,MODEP,JOEP	Année, mois, jour rela- tifs à la première observation	I2,1X,I2, 1X,I2	P1.FOR
5	ANFIN,MOFIN,JOFIN	Année, mois, jour rela- tifs à la dernière observation	I2,1X,I2, 1X,I2	P1.FOR
6	NOBS ICOD TN	Nombre d'observations =1 si l'utilisateur a déjà spécifié son propre titre de norme =0 Sinon Titre de la norme (si ICOD = 1)	I4,1X,I1,IX A15	P1.FOR P7.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT	ORIGINE
7	NMINABER	Nombre de valeurs aberrantes minimum éliminées.	I4,1X,I4,IX I4,1X,I4	P2.FOR
	NMAXABER	Nombre de valeurs aberrantes maximum éliminées.		
	LAVANT	Nombre de valeurs retirées au début de la série	P3.FOR	
	LAPRES	Nombre de valeurs retirées à la fin de la série		
8	ISAIS	=0 si on n'a pas détecté de saisonnalité =1 si on en a détecté une =2 si on a tronqué et qu'il n'y a pas de saisonnalité =3 si on a tronqué et qu'il y a saisonnalité	I1,1X,I2, 1X,I2,1X,I2	P3.FOR
	MD	Mois de départ, utilisé pour la redivision en saisons ou la troncation		
	INTER,NINT	Longueur de l'intervalle, en mois, et nombre d'intervalles par année (utilisés également pour la redivision en saisons ou la troncation)		
9	ICQ,AOR,BOR	ICQ=0 si la régression n'est pas significative ICQ=1 si elle l'est Dans l'affirmative, AOR et BOR sont les valeurs de a et b issues de la régression	I1,1X,F9.2, 1X,F9.2	P3.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT	ORIGINE
10	ITRONC IANTR IDURTR	=0 Faible fréquence identifiée par le logiciel mais non utilisée =1 Troncation =2 Moyennes saisonnières =3 Identifiée par l'usa- ger mais non traitée Première année identi- fiée comme ayant une faible fréquence d'échantillonnage Nombre d'années identi- fiées comme ayant une faible fréquence d'échantillonnage	I1,1X,I2, 1X, I2	P3.FOR et édité dans P4.FOR
11	INTERV IBIN NINTAN N	Intervalle choisi =0 en jours =1 en mois Nombre d'intervalles par années Longueur totale de la série, du 1er janvier de la 1ère année, jus- qu'au 31 décembre de la dernière année	I2,1X,I1,1X I3,1X,I4	P4.FOR
12	IEX1,IEX2 LS IDIV	Indices des extrémités non-absentes du vecteur (de longueur N, ci-haut mentionné) Longueur de la série avec les valeurs absen- tes retirées aux extré- mités on a $LS = IEX2 - IEX1 + 1$ Façon dont l'usager a finalement décidé de diviser sa série: =0 en mois/jours (division normale) =1 en reconstituant les saisons choisies dans P3 =2 en éliminant certains mois, tel que suggéré dans P3 (troncation)	I4,1X,I4,1X I4,1X,I1,1X I1	P4.FOR

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT	ORIGINE
	IBT	Méthode utilisée pour boucher les trous: =0 aucune =1 méthode des moyennes =2 interpolation =3 relation conc./débit		
13	IPERS,RHO	Résultats de l'étude de persistance: IPERS =0 pas de persistance significative IPERS =1 persistance détectée RHO : valeur de ρ s'il y a persistance significative (=0 sinon)	I1,1X,F5.2	P4.FOR
14	N1,N2 LONG	Extrémités désignées par l'usager pour le traitement (s'il choisit toute la série, N1=1 et N2=LS) Longueur de la série/ sous-série choisie On a $N2-N1+1 = LONG$	I4,1X,I4,1X I4	P5.FOR
15	NOSUGG NOTEST	Numéro du test suggéré par le logiciel. Numéro du test choisi par l'usager	I2,1X,I2	P5.FOR
16	STAT VALTEST SIGNIF COMMEN	Valeur de la statistique Valeur du test Niveau de signification du test Commentaire	E14.8,1X, E14.8,1X, F5.3,1X,A39	TESTS
17	IVAR	MW/MWS: représente la longueur du premier segment (avant le saut) en nombre d'intervalles à partir de N1 SP : nombre de degrés de liberté autres tests: - 9 (aucune signification)	I4,1X,I1	TESTS

LIGNE	NOM DES VARIABLES	DESCRIPTION	FORMAT	ORIGINE
	ITEND	=0 si aucune tendance n'a été détectée =1 si on a détecté une tendance		
18	RMSE	Racine carré de l'erreur quadratique moyenne pour le modèle de tendance choisi.	E16.10	P6.FOR

13.3 Autres_fichiers

Les fichiers de valeurs critiques ainsi que le fichier CODES.NAQ sont présentés dans cette section. Ces fichiers sont présentés afin que l'utilisateur ait une meilleure connaissance de certains traitements effectués dans le logiciel.

Seul le fichier CODES.NAQ peut être édité sans danger par l'utilisateur. Il peut alors ajouter des lignes sous le même format afin de compléter la gamme des codes reconnus. Des informations pourraient également être ajoutées en bout de ligne afin de préciser la signification du code analytique.

Il serait cependant très hasardeux d'éditer les fichiers de valeurs critiques puisqu'ils sont lus selon des formats très particuliers.

Tableau 13.3.1 Fichier STUDENT.TAB

COLONNE 1 = DEGRES DE LIBERTES,
COLONNE 2 = 95e PERCENTILE DE LA DISTRIBUTION DE STUDENT
COLONNE 3 = 97.5e PERCENTILE DE LA DISTRIBUTION DE STUDENT

1	6.31	12.7
2	2.92	4.30
3	2.35	3.18
4	2.13	2.78
5	2.01	2.57
6	1.94	2.57
7	1.90	2.36
8	1.86	2.31
9	1.83	2.26
10	1.81	2.23
11	1.80	2.20
12	1.78	2.18
13	1.77	2.16
14	1.76	2.14
15	1.75	2.13
16	1.75	2.12
17	1.74	2.11
18	1.73	2.10
19	1.73	2.09
20	1.72	2.09
21	1.72	2.08
22	1.72	2.07
23	1.71	2.07
24	1.71	2.06
25	1.71	2.06
26	1.70	2.06
27	1.70	2.05
28	1.70	2.05
29	1.70	2.04
30	1.70	2.04

Tableau 13.3.2 Fichier FISCHER.TAB

COLONNE 1 = DEGRES DE LIBERTE DU DENOMINATEUR
 COLONNE 2 A 11 = 95e PERCENTILE DE LA DISTRIBUTION DE FISCHER
 POUR 3,4,5,6,7,8,9,10,11 DEGRES DE LIBERTE AU
 NUMERATEUR.

1161.40	199.50	215.70	224.60	230.20	234.00	236.80	238.90	240.50	241.90	243.00
2 18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.40
3 10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76
4 7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93
5 6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70
6 5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03
7 5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60
8 5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31
9 5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10
10 4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94
11 4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82
12 4.75	3.88	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72
13 4.67	3.80	3.41	3.18	3.02	2.92	2.83	2.77	2.71	2.67	2.63
14 4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.56
15 4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51
16 4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.45
17 4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41
18 4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37
19 4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34
20 4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31
21 4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28
22 4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26
23 4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24
24 4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22
25 4.24	3.38	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20
26 4.22	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.18
27 4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.16
28 4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.15
29 4.18	3.33	2.93	2.70	2.54	2.43	2.35	2.28	2.22	2.18	2.14
30 4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.12
32 4.15	3.30	2.90	2.67	2.51	2.40	2.32	2.25	2.19	2.14	2.10
34 4.13	3.28	2.88	2.65	2.49	2.38	2.30	2.23	2.17	2.12	2.08
36 4.11	3.26	2.86	2.63	2.48	2.36	2.28	2.21	2.15	2.10	2.06
38 4.10	3.25	2.85	2.62	2.46	2.35	2.26	2.19	2.14	2.09	2.05
40 4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04
42 4.07	3.22	2.83	2.59	2.44	2.32	2.24	2.17	2.11	2.06	2.02
44 4.06	3.21	2.82	2.58	2.43	2.31	2.23	2.16	2.10	2.05	2.01
46 4.05	3.20	2.82	2.57	2.42	2.30	2.22	2.14	2.09	2.04	2.00
48 4.04	3.19	2.80	2.56	2.41	2.30	2.21	2.14	2.08	2.03	1.99
50 4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.02	1.98
55 4.02	3.17	2.78	2.54	2.38	2.27	2.18	2.11	2.05	2.00	1.97
60 4.00	3.15	2.76	2.52	2.37	2.25	2.17	2.10	2.04	1.99	1.95
65 3.99	3.14	2.75	2.51	2.36	2.24	2.15	2.08	2.02	1.98	1.94
70 3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.01	1.97	1.93
80 3.96	3.11	2.72	2.48	2.33	2.21	2.12	2.05	1.99	1.95	1.91
100 3.94	3.09	2.70	2.46	2.30	2.19	2.10	2.03	1.97	1.92	1.88
125 3.92	3.07	2.68	2.44	2.29	2.17	2.08	2.01	1.95	1.90	1.86
150 3.91	3.06	2.67	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85
200 3.89	3.04	2.65	2.41	2.26	2.14	2.05	1.98	1.92	1.87	1.83
400 3.86	3.02	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81
1000 3.85	3.00	2.61	2.38	2.22	2.10	2.02	1.95	1.89	1.84	1.80

Tableau 13.3.3 Fichier CHIDEUX.TAB

COLONNE 1 = DEGRES DE LIBERTE

COLONNE 2 = 95e PERCENTILE DE LA DISTRIBUTION DU CHIDEUX

0	0.00
1	3.84
2	5.99
3	7.81
4	9.49
5	11.1
6	12.6
7	14.1
8	15.5
9	16.9
10	18.3
11	19.7
12	21.0
13	22.4
14	23.7
15	25.0
16	26.3
17	27.6
18	28.9
19	30.1
20	31.4
21	32.7
22	33.9
23	35.2
24	36.4
25	37.7
26	38.9
27	40.1
28	41.3
29	42.6
30	43.8
40	55.8
50	67.5
60	79.1
70	90.5
80	101.9
90	113.14
100	124.34
200	250.00

Tableau 13.3.4 Fichier TESTRHO.TAB

COLONNE 1 = DEGRES DE LIBERTE
COLONNE 2 = VALEUR CRITIQUE AU SEUIL 5% DU TEST SUR LA SIGNIFICATION
DU COEFFICIENT DE CORRELATION
COLONNE 3 = VALEUR CRITIQUE AU SEUIL 2.5% DU TEST SUR LA SIGNIFICATION
DU COEFFICIENT DE CORRELATION

3	.878	.959
4	.811	.917
5	.754	.874
6	.707	.834
7	.666	.798
8	.632	.765
9	.602	.735
10	.576	.708
11	.553	.684
12	.532	.661
13	.514	.641
14	.497	.623
15	.482	.606
16	.468	.590
17	.456	.575
18	.444	.561
19	.433	.549
20	.423	.537
25	.381	.487
30	.349	.449
35	.325	.418
40	.304	.393
45	.288	.372
50	.273	.354
60	.250	.325
70	.232	.302
80	.217	.283
90	.205	.267
100	.195	.254
150	.160	.208
200	.138	.181
300	.113	.148
400	.098	.128
500	.088	.115

Tableau 13.3.5 Fichier FS2.TAB

COLONNE 1 = VALEUR DE LA PERSISTANCE
COLONNE 2 A 10 = ECART-TYPES OBTENUS A L'AIDE DE SIMULATIONS POUR
LES VALEURS DE N : 10, 25, 50, 75, 100, 150, 200, 300, 1000

.00	3.20	3.73	4.00	4.31	4.40	4.64	4.49	4.89	5.28
.10	3.36	3.85	4.22	4.27	4.45	4.72	4.65	4.82	5.56
.20	3.54	4.00	4.31	4.54	4.63	4.80	4.87	5.10	5.71
.30	3.65	4.10	4.44	4.73	4.79	5.20	5.16	5.34	5.67
.40	3.89	4.39	4.79	4.85	4.98	5.14	5.42	5.51	6.02
.50	4.13	4.68	4.99	5.14	5.08	5.34	5.30	5.44	6.12
.60	4.44	5.07	5.37	5.55	5.67	5.86	5.94	5.97	6.36
.70	4.91	5.71	6.08	6.26	6.35	6.05	6.59	6.88	6.61
.80	5.40	6.57	6.96	7.10	7.48	7.33	7.36	7.46	8.08
.90	6.04	8.45	9.23	9.54	9.52	9.88	9.92	9.95	10.48
.95	6.48	9.84	11.80	12.08	12.90	12.56	12.84	12.24	12.79

Tableau 13.3.6 Fichier CODES.NAQ

COLONNE 1 = CODES NAQUADAT
COLONNE 2 = EXTENSIONS CORRESPONDANTES POUR LES FICHIERS CREEES

00106 BID	
00120 SCA	27002 CO
00125 SAN	27301 CO
02011 COU	28002 NI
02021 COV	28302 NI
02041 CON	29005 CU
02051 CTA	29006 CU
02061 TEM	29305 CU
02073 TUR	30004 ZN
02074 TUR	30005 ZN
06001 COT	30305 ZN
06002 COT	33008 AS
06101 COD	34008 SN
06104 COD	48002 CD
06537 PHE	48301 CD
06901 COP	48302 CD
06902 COP	80011 HG
07016 NKJ	80311 HG
07010 NKJ	82002 FB
07112 AZO	82302 FB
07106 AZO	17811 HCB
07105 AZO	17814 HCB
07555 NH3	18000 DT1
08102 DD	18028 DT2
08101 DD	18005 DT3
09105 F	18027 DT4
10101 ALC	18010 TD1
10103 ALC	18013 TD2
10106 ALC	18020 DD1
10301 PH	18029 DD2
10401 SS	18030 MC1
10501 SS	18035 MC2
10603 DUR	18040 HC1
10602 DUR	18039 HC2
10606 DUR	18045 HC3
11101 NA	18044 HC4
11103 NA	18050 AE1
11105 NA	18085 AE2
12102 HG	18055 BE1
15406 PHO	18087 BE2
15413 PHO	18060 AC1
16301 SU	18059 AC2
16303 SU	18065 GC1
16306 SU	18064 GC2
17203 CL	18070 GB1
19102 K	18083 GB2
19103 K	18073 FB1
19105 K	18075 BHM
20003 CA	18077 BHM
20101 CA	18078 BB1
20103 CA	18125 MI1
36002 CT	18127 MI2
36013 CF	18130 AL1
36102 STF	18134 AL2
06604 CN	18140 EN1
06605 CN	18144 EN2
13003 AL	18150 DI1
13101 AL	18154 DI2
13302 AL	18164 ARO
13305 AL	18158 PCB
24003 CR	18415 AT1
24304 CR	18418 AT2
24305 CR	18419 AT3
25304 MN	18420 SIM
25305 MN	18430 CYP
25004 MN	18440 CVA
25005 MN	18500 24D
26004 FE	18510 24S
26005 FE	18540 FEN
26304 FE	
26305 FE	

14. Fiches techniques des tests

Nous présentons ici des fiches pour les différents tests afin de permettre à l'utilisateur d'avoir à la portée de la main une bonne source de renseignements sur les tests disponibles. De plus les tables de valeurs critiques qui sont citées sont présentées à la suite pour une consultation facile et rapide.

14.1 Notations

A) Études univariées

N: longueur de la série

n_1 : longueur de la première partie de la série; utilisé dans le cas des tendances par saut

n_2 : longueur de la deuxième partie de la série; utilisé dans le cas des tendances par saut

$\{x_1, \dots, x_N\}$: ensemble des réalisations de la série

$\{R_1, \dots, R_N\}$: ensemble des rangs des réalisations $x_1 \dots x_N$; sera une permutation de $\{1, \dots, N\}$

$\{1, \dots, N\}$: ensemble de référence pour les études de corrélation (Kendall, Spearman). Ici $1, \dots, N$ représentera le temps.

ρ_1 : coefficient d'autocorrélation d'ordre 1 de la série x_1, \dots, x_N .

B) Études avec saisonnalités

N : longueur de la série

S : nombre de saisons

n_s : nombre d'observations dans la saison "s"

n_{s1} : longueur de la première partie de la sous-série associée à la saison "s"; utilisé dans le cas de tendance par saut

n_{s2} : longueur de la deuxième partie de la sous-série associée à la saison "s"; utilisé dans le cas de tendance par saut

$\{x_1, \dots, x_N\}$: ensemble des réalisations de la série qui se divise de la façon suivante:

Années	Saisons				
	1	2	3	...	S
1	x_{11}	x_{12}	x_{13}	...	x_{1S}
2	x_{21}	x_{22}	x_{23}	...	x_{2S}
.
.
.
.
n	x_{n1}	x_{n2}	x_{n3}	...	x_{nS}

On note alors que si aucune donnée n'est manquante on a: $n_1 = n_2 = \dots = n_S = n$
 et que $\sum_{s=1}^S n_s = N$. De plus:

x_{ij} = observation associée à la saison "j" de l'année "i"

$\{R_1, \dots, R_N\}$: ensemble des rangs associés aux réalisations $\{x_1, \dots, x_N\}$

avec:

R_{ij} = rang associé à la valeur pour la saison "j" de l'année "i"

N.B. . À moins d'information contraire, les rangs sont attribués à l'intérieur de chaque saison. On aura donc des permutations des ensembles $\{1, \dots, n_s\}$, $s = 1 \dots S$

. Dans le cas de la détection d'une tendance par saut, la séparation sera sur une année complète. Cette année étant choisie par une date ou un rang. Toutes les saisons seront donc divisées à la même année.

C) Autres notations importantes

H_α : valeur critique au seuil $\alpha\%$

Z_α : α^e percentile de la loi Normale centrée-réduite

$t_{\alpha,n}$: α^e percentile de la loi de Student à n degrés de liberté

$x_{\alpha,n}^2$: α^e percentile de la loi de Chi-Deux à n degrés de liberté

14.2 Application générale des tests

On suppose généralement la symétrie de H_α par rapport à 0, i.e. $H_{(1-\alpha)} = -H_\alpha$.
Dans un cas contraire, on présentera à la fois $H_{(1-\alpha)}$ et H_α .

Ainsi pour un test unilatéral voulant détecter une tendance négative (diminution) on utilise la statistique du test T, pour construire le système de décision suivant:

Rejet de H_0 Si $T \leq -H_\alpha$

Acceptation de H_0 Si $T > -H_\alpha$

Test unilatéral voulant détecter une tendance positive

Rejet de H_0 Si $T \geq H_\alpha$

Acceptation de H_0 Si $T < H_\alpha$

Test bilatéral voulant détecter une tendance positive ou négative

Rejet de H_0 Si $|T| \geq H_{\alpha/2}$

Acceptation de H_0 Si $|T| < H_{\alpha/2}$

N.B. Tous ces tests sont effectués au seuil $\alpha\%$

14.3 Glossaire des termes techniques présents dans les fiches

- E.A.R. Efficacité Asymptotique Relative. Équivaut au rapport des grandeurs d'échantillons nécessaire à deux tests pour avoir la même puissance sous une loi F.
- MARKOVIEEN Relatif à des structures particulières de persistance. Un processus markovien d'ordre 1 à la structure $\rho_k = \rho_1^k$ où ρ_k = coefficient d'autocorrélation d'ordre k.
- PERSISTANCE Représente une dépendance des données, une tendance à court terme. Un ensemble de données (échantillon) est persistant si les données successives sont corrélées, on voit donc un rapprochement certain avec l'autocorrélation.
- SÉRIES SYNCHRONES: Se dit de 2 (ou plus) séries qui sont échantillonnées aux mêmes intervalles de temps.
- TENDANCE MONOTONE: Consiste en un changement graduel de la valeur d'un paramètre dans une direction unique. Exemple: la moyenne de la concentration de matières en suspension augmente au rythme de 1 mg/l par année.
- TENDANCE PAR SAUT: Consiste en un changement instantané de la valeur d'un paramètre. Exemple: la moyenne de la concentration de matières en suspension est passé de 3 mg/l à 6 mg/l à une date fixe suite à un rejet industriel.

14.4 Structure des fiches de test

OBJET DU TEST: 1-Hypothèse nulle. 2-Hypothèse alternative.

CONDITIONS D'APPLICATION: Conditions dans lesquelles le test doit être effectué. Exemple: indépendance, normalité, etc.

APPLICATION POUR LES TENDANCES: Explication des changements à effectuer pour utiliser ce test dans le but de détecter des tendances.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Renseignements concernant l'application en hydrologie et renvoie à un test plus approprié lorsque les caractéristiques particulières entrent en conflit avec les conditions d'application.

EXPRESSION ET MODE DE CALCUL: Présentation de la statistique utilisée pour le test et de la façon dont on l'obtient.

VALEURS CRITIQUES: 1-Présentation des valeurs critiques pour petits et grands échantillons. 2-Tableau sommaire des valeurs critiques pour un seuil de 95%.

CORRECTIONS: 1-Correction de continuité. 2-Corrections en présence de valeurs (rangs) égales.

PUISSANCE: Présentation des résultats connus quant à la puissance et à la robustesse du test.

RÉFÉRENCES: 1-Référence statistique. 2-Référence sur l'application dans le domaine de la qualité de l'eau.

14.5 Fiches des tests

Les fichiers suivants donnant un résumé des hypothèses, modes de calcul et valeurs critiques de tests utilisés.

TEST DE MANN WHITNEY

FICHE # 1

OBJET DU TEST: 1- Égalité des moyennes de deux échantillons ($\mu_1 = \mu_2$). 2- Moyennes différentes, sous forme unilatérale ($\mu_1 \leq \mu_2$ ou $\mu_2 \leq \mu_1$) ou bilatérale ($\mu_1 \neq \mu_2$).

CONDITIONS D'APPLICATION: La série de mesures ne doit pas contenir de persistance ou de saisonnalité.

APPLICATION POUR LES TENDANCES: On divise la série en deux parties 1, ..., n_1 et $(n_1 + 1)$, ..., N, avec $n_2 = N - n_1$, et n_2 n'a pas à égaliser n_1 . On peut alors tester pour détecter une tendance sur la moyenne. Ce test est plus approprié pour détecter une tendance par saut, mais il détectera également des tendances monotones.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Dans la cas où une persistance est présente, on doit se référer au test de LETTENMAIER MANN-WHITNEY, fiche # 2. Si une composante de saisonnalité est présente, on peut alors utilisé le test de MANN-WHITNEY saisonnier fiche # 3.

EXPRESSION ET MODE DE CALCUL: On donne les rangs pour l'échantillon complet $x_1 \dots x_N$. Soit R_{i1} = le rang de la i^e observation du groupe 1. On utilise alors la statistique:
$$U = \sum_{i=1}^{n_1} R_{i1}$$

VALEURS CRITIQUES: Pour de petits échantillons, des tables sont présentées dans CONOVER (1971 table A7 p. 448). Pour les grands échantillons ($n_1 \geq 20$, $n_2 \geq 20$ et $n_2 \geq n_1$), on utilise une approximation normale avec:
 $E(U) = n_1(N+1)/2$, $VAR(U) = n_1 n_2 (N+1)/12$

et
$$H_\alpha = E(U) + Z_\alpha VAR(U)^{\frac{1}{2}}$$

Tableau sommaire pour un seuil de 95%.

	$n_1 = n_2 = 20$	$n_1 = 20,$ $n_2 = 50$	$n_1 = n_2 = 50$	$n_1 = 50$ $n_2 = 100$
unilatéral	471	902	2764	7963
bilatéral	(338,482)	(557,928)	(2240,2810)	(3283,8042)

CORRECTIONS: 1. On utilise la correction: $H_\alpha^* = H_\alpha - \frac{1}{2}$ lorsque l'on doit travailler avec l'approximation normale. 2. On utilise les rangs moyens et on ajuste la variance de la façon suivante:

$$VAR(U) = \frac{12}{n_1 n_2 (N+1)} \left[1 - \sum_{s=1}^S \frac{t_s^3 - t_s}{N_s - N} \right]$$

où s = nombre de groupes où il y a égalité et où t_s = nombre de valeurs égales dans le groupe s .

PUISSANCE: On utilise souvent ce test à cause de sa puissance: EAR = 0.955 versus le test de Student pour des données normales et EAR ≥ 1.0 pour plusieurs distributions autres que normale.

RÉFÉRENCES: 1- LEHMAN, E.L. (1975). Nonparametrics: Statistical Methods Based on Ranks. Holden-Day, San Francisco, 457 p.

TEST DE LETTENMAIER/MANN-WHITNEY

FICHE # 2

OBJET DU TEST: 1- Égalité des moyennes de deux échantillons ($\mu_1 = \mu_2$). 2- Moyennes différentes, sous forme unilatérale ($\mu_1 \leq \mu_2$ ou $\mu_2 \leq \mu_1$) ou bilatérale ($\mu_1 \neq \mu_2$).

CONDITIONS D'APPLICATION: la série contient une persistance, à court terme, de type markovien. La série ne doit pas contenir de saisonnalités.

APPLICATION POUR LES TENDANCES: On divise la série en deux parties 1, ..., n_1 et (n_1+1) , ..., N, avec $n_2 = N - n_1$, et n_2 n'a pas à égaier n_1 . On peut alors tester pour détecter une tendance sur la moyenne. Ce test est plus approprié pour détecter une tendance par saut, mais il détectera également des tendances monotones.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Si une composante de saisonnalité est présente, on peut alors utilisé le test de MANN-WHITNEY SAISONNIER / LETTENMAIER fiche # 4. On calcule n_1^* et n_2^* à l'aide du I obtenu avec N.

EXPRESSION ET MODE DE CALCUL: On donne les rangs pour l'échantillon complet $x_1 \dots x_N$. Soit R_{i1} = le rang de la i^e observation du groupe 1. On utilise de nouveau la statistique:
$$U = \sum_{i=1}^{n_1} R_{i1}$$

VALEURS CRITIQUES: On utilise les mêmes résultats que pour le test de MANN-WHITNEY en ajustant la variance à l'aide du contenu en information, I (voir fiche # A 1).

$$E(U) = n_1(N+1)/2, \text{ VAR}(U) = n_1 n_2 (N+1)/(12.I)$$

et
$$H_\alpha = E(U) + Z_\alpha \text{VAR}(U)^{\frac{1}{2}}$$

CORRECTIONS: 1. On utilise la correction: $H_\alpha^* = H_\alpha - \frac{1}{2}$ lorsque l'on travaille avec l'approximation normale. 2. On utilise les rangs moyens et on ajuste la variance de la façon suivante:

$$\text{VAR}(U) = \frac{n_1 n_2 (N+1)}{12} \left[1 - \sum_{s=1}^S \frac{t_s^3 - t_s}{N_3 - N} \right]$$

où s = nombre de groupe où il y a égalité et où t_s = nombre de valeurs égales dans le groupe s .

PUISSANCE: La puissance de ce test est équivalente à celle de MANN-WHITNEY pour le N^* .

RÉFÉRENCES: 2- LETTENMAIER, D.P. (1976). Detection of Trends in Water Quality Data From Records with Dependent Observations. Water Resources Research. 12(5) 1037-1046.

TEST DE MANN-WHITNEY SAISONNIER

FICHE # 3

OBJET DU TEST: 1- Égalité des moyennes de deux échantillons contenant des saisonnalités. 2- Moyennes différentes.

CONDITIONS D'APPLICATION: la série de mesures ne doit pas contenir de persistance. Ce test n'est pas recommandé pour les cas où l'on a pas 6 années d'observations par saison.

APPLICATION POUR LES TENDANCES: On divise les groupes d'observations de chaque saison, en deux parties qui n'ont pas à être égales. Les séparations doivent être consistantes d'une saison à l'autre.

RENSEIGNEMENTS SUPPLÉMENTAIRES: En présence de persistance, on doit se référer au test de MANN-WHITNEY SAISONNIER/LETTENMAIER fiche # 4.

EXPRESSION ET MODE DE CALCUL: On calcule à l'intérieur de chaque saison une statistique de MANN-WHITNEY, fiche # 1, que l'on appelle U_s . On utilise alors pour ce test: $US = \sum_{s=1}^S U_s$.

VALEURS CRITIQUES: On utilise les résultats théoriques suivants:

$$E(US) = \sum_{s=1}^S E(U_s) \text{ et } \text{VAR}(US) = \sum_{s=1}^S \text{VAR}(U_s).$$

pour construire:

$$H_{\alpha} = E(US) + Z_{\alpha} (\text{VAR}(US))^{0.5}.$$

CORRECTIONS: 1. On utilise la correction: $H_{\alpha}^* = H_{\alpha} - \frac{1}{2}$ lorsque l'on travaille avec l'approximation normale. 2. On travaille à l'intérieur de chaque saison comme pour le test de MANN-WHITNEY.

PUISSANCE: Aucune étude de puissance connue.

RÉFÉRENCES: 2- BERRYMAN, D. (1984). La détection de tendance dans les séries temporelles de paramètres de la qualité de l'eau à l'aide de tests non paramétriques. Mémoire de maîtrise déposé à l'INRS-Eau.

TEST DE MANN-WHITNEY SAISONNIER/LETTENMAIER

FICHE # 4

OBJET DU TEST: 1- Égalité des moyennes de deux échantillons contenant des saisonnalités. 2- Moyennes différentes.

CONDITIONS D'APPLICATION: Comme la puissance de ce test n'est pas connue, il est préférable d'avoir un échantillon assez grand pour chaque saison. Ainsi, pour les cas où l'on n'a pas 6 années d'observation par saison, ce test n'est pas recommandé.

APPLICATION POUR LES TENDANCES: On divise les groupes d'observations de chaque saison en deux parties qui n'ont pas à être égales. Les séparations doivent être consistantes d'une saison à l'autre.

RENSEIGNEMENTS SUPPLÉMENTAIRES: On calculera tous les nombres effectifs à l'aide du Contenu en Information I, voir fiche # 2.

EXPRESSION ET MODE DE CALCUL: On calcule à l'intérieur de chaque saison une statistique de MANN-WHITNEY, fiche # 1, que l'on appelle U_s . On utilise alors pour ce test: $US = \sum_{s=1}^S U_s$.

VALEURS CRITIQUES: On utilise les mêmes valeurs théoriques:

$$E(US) = \sum_{s=1}^S \frac{E(U_s)}{S} \quad \text{et} \quad \text{VAR}(US) = \sum_{s=1}^S \text{VAR}(U_s).$$

pour construire:

$$H_\alpha = E(US) + Z_\alpha \text{VAR}(US)^{0.5}.$$

N.B. Les VAR (U_s) sont calculées avec le contenu en information I obtenu à partir de N.

CORRECTIONS: 1. On utilise la correction: $H_\alpha^* = H_\alpha - \frac{1}{2}$. 2- On travaille à l'intérieur de chaque saison comme pour le test MANN-WHITNEY.

PUISSANCE: Aucune étude de puissance connue.

RÉFÉRENCES: 2- BERRYMAN, D. (1984). La détection de tendance dans les séries temporelles de paramètres de la qualité de l'eau à l'aide de tests non paramétriques. Mémoire de maîtrise déposé à l'INRS-Eau.

TEST DE KENDALL

FICHE # 5.1

OBJET DU TEST: 1-Indépendance entre deux séries de mesures. 2-Tendances monotones, négatives ou positives, sous forme unilatérale ou bilatérale.

CONDITIONS D'APPLICATION: La série ne doit pas contenir de saisonnalités ou de persistance

APPLICATION POUR LES TENDANCES: Application directe en utilisant le temps comme une série de référence.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Ce test apparaît bien approprié pour détecter une tendance monotone linéaire, mais peut aussi être utilisé pour détecter d'autres types de tendances monotones. En présence de saisonnalités, on doit se référer au test de KENDALL SAISONNIER fiche # 7 et en présence de persistance, on utilisera le test de LETTENMAIER/SPEARMAN fiche # 6.

EXPRESSION ET MODE DE CALCUL: Soit P, le nombre de paires concordantes et Q, le nombre de paires discordantes parmi les $N(N-1)/2$ paires distinctes possibles ($P+Q = N(N-1)/2$). Une paire est dite concordante si pour $i \leq j$ alors $R_i \leq R_j$, inversement, une paire est dite discordante si pour $i \leq j$ alors $R_i \geq R_j$. On utilise alors la statistique: $K = P - Q$.

VALEURS CRITIQUES: 1-Pour $N \leq 60$, CONOVER (1971) présente les valeurs critiques pour différents seuils. Alors que pour $N \geq 60$ on utilise l'approximation du percentile de H par: $H_\alpha = Z_\alpha [(N(N-1)(2N+5))/18]^{0.5}$.

2-Tableau sommaire des valeurs critiques pour un seuil de 95%.

N	4	5	6	7	10	15	25	40	60	100	150
unilatéral	4	6	9	11	19	33	70	142	258	553	1013
bilatéral	6	8	11	13	21	39	84	168	306	659	1206

CORRECTIONS: On utilise la correction de continuité $H_\alpha^* = H_\alpha + 1$ avec un rejet pour $K > H_\alpha^*$ et non $K \geq H_\alpha^*$. 2-On attribue le rang moyen à chacune des valeurs égales et on ajuste H_α :

$$H_\alpha = Z_\alpha \left[N(N-1)(2N+5)/18 - \sum_{a=1}^A t_a(t_a-1)(2t_a+5)/18 \right]^{0.5}$$

où A = nombre de groupes de valeurs égales et où t_a = nombre de valeurs égales pour le groupe a.

PUISSANCE: Si (X_i, Y_i) , $i=1 \dots N$ sont indépendants et identiquement distribués selon une loi bivariée, on obtient alors les E.A.R. suivantes pour K par rapport à l'utilisation du r de Pearson (paramétrique) comme statistique d'un test: 0.912 en présence d'une loi normale, 1.0 en présence d'une loi uniforme et 1.266 en présence d'une loi double exponentielle (STUART; 1954).

RÉFÉRENCES: 1- CONOVER, W.J. (1971). Practical Nonparametric Statistics. 2ed. John Wiley, New York, 493 p. 2- HIRSCH, R.M. Jr. et SLACK, J.K. (1982). Techniques of Trend Analysis for Monthly Water Quality Data. Water Resources Research, 18(1): 107-121.

TEST DU TAU DE KENDALL

FICHE # 5.2

OBJET DU TEST: 1- Indépendance entre deux séries de mesures synchrones.
2- Tendances monotones, négative ou positive, sous forme unilatérale ou bilatérale.

CONDITIONS D'APPLICATION: La série de mesures ne doit pas contenir de saisonnalités ou de persistance.

APPLICATION POUR LES TENDANCES: Application directe en utilisant le temps comme une des séries de mesures.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Ce test semble plus approprié pour détecter une tendance monotone linéaire mais son utilisation peut également permettre de détecter d'autres types de tendances monotones. En présence de saisonnalités on doit se reporter au test de KENDALL SAISONNIER, fiche # 7 et en présence de persistance on utilisera le test de LETTENMAIER/SPEARMAN fiche # 6.

EXPRESSION ET MODE DE CALCUL: On calcule d'abord K comme pour le test de KENDALL mais on utilise plutôt ici la statistique $\tau = \frac{2K}{N(N-1)}$. On a alors τ variant entre -1 et 1 ce qui permet un rapprochement avec les autres mesures de corrélation (r de PEARSON et rho de SPEARMAN). Il faut cependant noter que pour un même échantillon de deux séries, on aura des valeurs différentes pour ces trois statistiques (τ de KENDALL, r de PEARSON et rho de SPEARMAN).

VALEURS CRITIQUES: Pour de petits échantillons, on peut utiliser la table A12 de CONOVER en divisant les valeurs critiques de cette table par $N(N-1)/2$. Pour de grands échantillons on utilise l'approximation suivante du percentile de H: $H_\alpha = Z_\alpha \left[\frac{2(2N+5)}{9N(N-1)} \right]^{0.5}$.

2- Tableau sommaire des valeurs critiques pour un seuil de 95%

N	4	5	6	7	10	15	25	40	60	100	150
unilatéral	0.67	0.60	0.60	0.52	0.42	0.31	0.23	0.18	0.15	0.11	0.09
bilatéral	1.00	0.80	0.73	0.62	0.47	0.37	0.28	0.22	0.17	0.13	0.11

CORRECTIONS: 1- Voir test de KENDALL fiche # 5.1. On attribue le rang moyen à chacune des valeurs égales et on ajuste H_α :

$$H_\alpha = Z_\alpha \left[\frac{2(2N+5)}{9(N(N-1))} - \sum_{a=1}^A \frac{2(2t_a+5)}{9(t_a(t_a-1))} \right]^{0.5} \quad \text{où } A \text{ et } a \text{ sont similaires au}$$

cas du test de KENDALL fiche # 5.1

PUISSANCE: Voir test de KENDALL fiche # 5.1

RÉFÉRENCES: Voir test de KENDALL fiche # 5.1

TEST DE SPEARMAN

FICHE # 5.3

OBJET DU TEST: 1- Indépendance entre deux séries de mesures synchrones.
2- Tendances monotones, négatives ou positives, sous forme unilatérale ou bilatérale.

CONDITIONS D'APPLICATION: Les séries de mesures ne doivent pas contenir de saisonnalités ou de persistance.

APPLICATION POUR LES TENDANCES: Application directe en utilisant le temps comme une des séries de mesures. L'autre série sera la série de mesures pour laquelle on veut détecter une tendance.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Ce test apparaît plus approprié pour détecter une tendance monotone linéaire, mais peut aussi être utilisé pour détecter d'autres types de tendances monotones. En présence de saisonnalités, aucun développement particulier n'a encore été construit. Étant donné la similitude, on peut cependant se reporter au test de KENDALL saisonnier, fiche # 7. En présence de persistance on utilisera le test de LETTENMAIER/SPEARMAN fiche # 6.

EXPRESSION ET MODE DE CALCUL: On utilise ici les rangs des observations pour construire la statistique. Soit R_i le rang de l'observation échantillonnée au temps de rang i . On pose alors $d_i = R_i - i$: la différence entre le rang de l'observation et le rang de la mesure dans le temps. On aura donc $R_i \in \{1, \dots, N\}$ et $i \in \{1, \dots, N\}$. La statistique utilisée est alors:

$$r_s = 1 - \frac{6 \sum_{i=1}^N d_i^2}{N^3 - N}$$

VALEURS CRITIQUES: On a toujours $|r_s| \leq 1$ et $r_s = 1$ si et seulement si tous les rangs des observations sont égaux à leur rang chronologique. Pour les petits échantillons $n \leq 11$, MOSTELLER et ROURKE présente une table des valeurs critiques pour des seuils près de 97.5%, 95% et 90%. Pour des grands échantillons, on utilise l'approximation $H_\alpha = [t_\alpha / (n-2 + t_\alpha)]^{0.5}$.

2- Tableau sommaire des valeurs critiques pour un seuil de 95% unilatéral

N	5	7	10	15	20	30	50	100
unilatéral	0.9	0.68	0.56	0.44	0.38	0.31	0.24	0.17
bilatéral	1	0.79	0.65	0.51	0.45	0.36	0.28	0.20

CORRECTIONS: 1- Aucune correction n'est présentée dans les références étudiées. 2- On attribue un rang moyen à chacune des valeurs et on ajuste

$$r_s: r_s = \frac{(N^3-N) - 6 \sum_{i=1}^N d_i^2 - 6(t' + u')}{\{[(N^3-N) - 1 t'] [(N^3-N) - 12 u']\}^{\frac{1}{2}}} \quad \text{où } t' = \sum_{a=1}^A \frac{t_a^3 - t_a}{12} \text{ et}$$

$$u' = \sum_{b=1}^B \frac{(u_b^3 - u_b)}{12}$$

A = nombre de cas d'égalités pour la série R;
 t_a = nombre de valeurs égales dans le groupe a;
 B = nombre de cas d'égalités pour la série i;
 u_b = nombre de valeurs égales pour le groupe b.

PUISSANCE: Tout comme le test de KENDALL, l'EAR du test de SPEARMAN est de 0.912 par rapport à l'utilisation du r de PEARSON comme statistique d'un test, CONOVER (1971). Donc la puissance de ce test est bonne.

RÉFÉRENCES: 1- MOSTELLER, F. et R.E.K. ROURKE (1973). Sturdy Statistics. Addison-Wesley, Reading. 395 p. 2- CONOVER, W.J. (1971). Practical Nonparametrics Statistics. 2éd. John Wiley et Sons, New York, 493 p.

TEST DE LETTENMAIER/SPEARMAN

FICHE # 6

OBJET DU TEST: 1- Indépendance entre deux séries de mesures synchrones.
2- Tendances monotones, négatives ou positives, sous forme unilatérale ou bilatérale.

CONDITIONS D'APPLICATION: Les séries de mesures ne doivent pas contenir de saisonnalités.

APPLICATION POUR LES TENDANCES: Application directe en utilisant le temps comme une des séries de mesures. L'autre série sera la série de mesures pour laquelle on veut détecter une tendance.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Ce test apparaît plus approprié pour détecter une tendance monotone linéaire, mais peut aussi être utilisé pour détecter d'autres types de tendances monotones. En présence de saisonnalités, aucun développement particulier n'a encore été construit, on doit alors se référer au test de HIRSCH et SLACK fiche # 8.

EXPRESSION ET MODE DE CALCUL: On utilise ici les rangs des observations pour construire la statistique. Soit R_i le rang de l'observation échantillonnée au temps de rang i . On pose alors $d_i = R_i - i$: la différence entre le rang de l'observation et le rang de la mesure dans le temps. On aura donc $R_i \in \{1, \dots, N\}$ et $i \in \{1, \dots, N\}$. La statistique utilisée est alors:

$$r_s = 1 - \left(6 \sum_{i=1}^N d_i^2 / (N^3 - N) \right).$$

VALEURS CRITIQUES: On utilise les mêmes valeurs critiques que pour le test de SPEARMAN, sauf que l'on travaille avec le nombre effectif N^* plutôt qu'avec N , voir fiche # A1.

CORRECTIONS: On travaille comme pour le test de SPEARMAN mais avec le nombre effectif N^* .

PUISSANCE: Même puissance que le test de SPEARMAN par rapport à l'utilisation du r de PEARSON, au niveau échantillonnal N^* .

RÉFÉRENCES: 2- LETTENMAIER, D.P. (1976). Detection of Trends in Water Quality Data from Records with Dependent Observations. Water Resources Research, 12(5), 1037-1046.

TEST DE KENDALL SAISONNIER

FICHE # 7

OBJET DU TEST: 1- Indépendance entre deux séries synchrones contenant des saisonnalités. 2- Tendances monotones, négatives ou positives, sous forme unilatérale ou bilatérale.

CONDITIONS D'APPLICATION: Les séries de mesures ne doivent pas comporter de dépendance autre que la saisonnalité et la tendance. On doit donc s'assurer qu'aucune persistance n'est présente.

APPLICATION POUR LES TENDANCES: Application directe pour la détection de tendance sur la moyenne en présence de saisonnalités, en utilisant le temps comme une des séries de mesures.

RENSEIGNEMENTS SUPPLÉMENTAIRES: En présence de persistance, on peut se reporter au test de HIRSCH et SLACK, fiche # 8.

EXPRESSION ET MODE DE CALCUL: Ce test utilise les statistiques de KENDALL calculées à l'intérieur de chaque saison. On divise d'abord la série de mesures en sections afin d'avoir, à l'intérieur de chaque section, des données qui semblent être affectées de façon homogène par la composante saisonnière. Soit S = nombre de saisons, on calcule pour $s = 1, \dots, S$: $k_s = P_s - Q_s$ où P_s = nombre de paires concordantes pour la saison s et où Q_s = nombre de paires discordantes pour la saison s (réf. test de KENDALL fiche # 5.1). Le nombre de données à l'intérieur de chaque saison n'a pas à être le même. La statistique utilisé est alors: $KS = \sum_{s=1}^S k_s$.

VALEURS CRITIQUES: Pour $n_s \geq 3$ $s = 1 \dots S$, on utilise les résultats:

$E(KS) = 0$ et $Var(KS) = \sum_{s=1}^S [n_s(n_s-1)(2n_s+5)]/18$ pour construire la statistique $\frac{KS}{(Var(KS))^{1/2}}$ qui suit approximativement une loi normale centrée-réduite.

On a donc: $H_\alpha = Z_\alpha \left[\sum_{s=1}^S [n_s(n_s-1)(2n_s+5)]/18 \right]^{0.5}$

Comme aucune table n'est disponible pour de petits échantillons ($n_s \leq 2$) et comme l'approximation normale n'est pas très précise dans ce cas, ce test ne sera pas utilisé dans de telles circonstance.

CORRECTIONS: 1- On utilise la correction $H_\alpha^* = H_\alpha + 1$. 2- On effectue à l'intérieur de chaque saison la correction présentée dans le test de KENDALL fiche # 5.1.

PUISSANCE: Pas de référence connue sur la puissance de ce test.

RÉFÉRENCES: 2- HIRSCH, R.M. et J.R. SLACK (1982). Techniques of Trend Analyses for Monthly Water Quality Data. Water Resources Research, 18(1): 107-121.

On utilise ce test pour compléter l'analyse à l'aide du test KENDALL SAISONNIER (fiche # 7). Ainsi, lorsque la tendance n'est pas homogène entre les saisons, le test de KENDALL SAISONNIER peut ne pas détecter de tendance générale malgré des tendances très évidentes à l'intérieur des saisons, pour un exemple particulier de ce résultat, voir VAN BELLE et HUGHES, 1984 p. 130.

Ce test est utilisé automatiquement par le logiciel lorsque l'on accepte l'absence de tendance à l'aide du test KENDALL SAISONNIER, on cherche alors à détecter des tendances particulières à l'intérieur des saisons.

Construction du test: On calcule les k_s comme dans le test KENDALL SAISONNIER et on construit ensuite:

$$\chi^2_{\text{total}} = \sum_{s=1}^S k_s^2 / \text{Var}(k_s) \text{ et } \chi^2_{\text{tend.}} = (KS)^2 / \text{Var}(KS)$$

on a finalement:

$$\chi^2_{\text{homogène}} = \chi^2_{\text{total}} - \chi^2_{\text{tend.}}$$

Sous l'hypothèse d'absence de tendance on a:

$$\chi^2_{\text{homogène}} \sim \chi^2_{s-1} \text{ et } \chi^2_{\text{tend.}} \sim \chi^2_1$$

Comme on a déjà accepté l'absence d'une tendance générale, le $\chi^2_{\text{tend.}}$ ne devrait pas être significatif. On teste ensuite afin de savoir si le $\chi^2_{\text{homogène}}$ est significatif. Dans l'affirmative, on conclut à la présence de tendances à l'intérieur des saisons et on recherche alors le "pattern" à cette hétérogénéité à l'aide des contrastes (VAN BELLE et HUGHES 1984, p. 130).

RÉFÉRENCE: VAN BELLE, G. et J.P. HUGHES (1984). Non Parametric Tests for Trends in Water Quality. Water Resources Research. 20.1, 127-136.

TEST DE HIRSCH ET SLACK

FICHE # 8

OBJET DU TEST: 1- Indépendance entre deux séries synchrones contenant des saisonnalités. 2- Tendances monotones, négatives ou positives, sous forme unilatérale ou bilatérale.

CONDITIONS D'APPLICATION: La série de mesures peut contenir persistance et saisonnalités.

APPLICATION POUR LES TENDANCES: Application directe pour la détection de tendance sur la moyenne en présence de saisonnalité et de persistance en utilisant le temps comme une des séries de mesures.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Ce test sera moins puissant que le test de KENDALL SAISONNIER fiche # 7 s'il n'y a pas de persistance.

EXPRESSION ET MODE DE CALCUL: On construit la statistique KS comme pour le test KENDALL SAISONNIER fiche # 7 et on la nomme HS.

VALEURS CRITIQUES: Comme les données ne sont pas indépendantes:

$$\text{VAR (HS)} = \sum_{s=1}^S \text{VAR} (k_s) + 2 \sum_{s=1}^S \sum_{l=s}^S \text{Cov} (k_s, k_l)$$

et la somme des termes de covariance n'est pas nulle.

On estime alors $\text{Cov} (k_s, k_l)$ avec:

$$\sigma_{s1} = [K_{s1} + 4 \sum_{i=1}^n R_{is} R_{il} - n(n+1)^2] / 3$$

$$\text{avec } K_{s1} = \sum_{i \leq j} \text{sign} [(x_{js} - x_{is}) (x_{j1} - x_{i1})]$$

et on suppose ici $n_s = n$ pour $s = 1 \dots S$.

x_{ij} = valeur mesurée pour l'année i et la saison j

R_{ij} = rang de x_{ij} à l'intérieur du groupe de valeurs de la saison j

Lorsque les n_s ne sont pas tous égaux, on utilise:

$$\sigma_{s1} = [K_{s1} + 4 \sum_{i=1}^n R_{is} R_{il} - n (n_s + 1) (n_1 + 1)] / 3.$$

On utilise finalement: $H_\alpha = Z_\alpha \cdot (\text{Var (HS)})^{0.5}$ avec Var (HS) calculée à l'aide des σ_{s1} .

CORRECTIONS: 1- On utilise la correction $H_\alpha^* = H_\alpha + 1$. 2- Les termes $\text{Var} (k_s)$ sont ajustés comme pour le test KENDALL SAISONNIER fiche # 7. Pour ce qui est des termes σ_{s1} ils demeurent les mêmes, mais on utilise les rangs moyens pour les cas d'égalité.

PUISSANCE: Aucune étude de puissance ne nous est connue. On peut cependant dire que l'approximation normale que l'on utilise sera bonne pour un ensemble ayant au moins une dizaine d'années.

RÉFÉRENCES: HIRSCH, R.M. et J.R. SLACK (1984). A nonparametric Trend Test for Seasonal Data With Serial Dependence. Water Resources Research, 20.6, 727-732.

TEST DE FOSTER ET STUART 1

FICHE # 9

OBJET DU TEST: 1- Indépendance des observations d'une série. 2- Tendances monotones négatives, positives (test unilatéral) et négatives ou positives (test bilatéral).

CONDITIONS D'APPLICATION: Afin de vraiment détecter une tendance, la série doit être exempte des autres types de dépendance (persistance et saisonnalités).

APPLICATION POUR LES TENDANCES: Application directe pour la détection de tendances monotones sur la moyenne.

RENSEIGNEMENTS SUPPLÉMENTAIRES: En présence de persistance, on utilise le test de FOSTER et STUART 2, fiche # 10.

EXPRESSION ET MODE DE CALCUL: On calcule d'abord pour $i = 2 \dots N$

$$m_i = \begin{cases} 1 & \text{Si } x_i \leq x_j \quad j \leq i \\ 0 & \text{sinon} \end{cases}, \quad M_i = \begin{cases} 1 & \text{Si } x_i \geq x_j \quad j \leq i \\ 0 & \text{sinon} \end{cases}, \quad d = \sum_{i=2}^N M_i - m_i$$

et pour $i = 1 \dots (N-1)$

$$m'_i = \begin{cases} 1 & \text{Si } x_i \leq x_j \quad i \leq j \\ 0 & \text{sinon} \end{cases}, \quad M'_i = \begin{cases} 1 & \text{Si } x_i \geq x_j \quad i \leq j \\ 0 & \text{sinon} \end{cases}, \quad d' = \sum_{i=1}^{N-1} M'_i - m'_i$$

la statistique utilisée est alors $D = d - d'$.

VALEURS CRITIQUES: 1- Pour $8 \leq N \leq 130$ et pour $\frac{1}{\sqrt{10N}} \leq \alpha \leq 1 - \frac{1}{\sqrt{10N}}$,

$$H_\alpha = (.48 + 2.78 \sqrt{\log_{10} N}) Z_\alpha. \quad \text{Pour } N \geq 130 \text{ et pour } \frac{1}{\sqrt{10N}} \leq \alpha \leq 1 - \frac{1}{\sqrt{10N}},$$

$$H_\alpha = \frac{1}{2} + [2(2 \ln N - 0.846)]^{0.5} Z_\alpha.$$

2- Tableau sommaire pour un seuil de 95%

N	10	25	50	75	100	125	150	200
unilatéral	5.4	6.2	6.7	7.1	7.3	7.4	7.5	7.8
bilatéral	6.4	7.4	8.1	8.4	8.6	8.8	8.9	9.2

CORRECTIONS: 1- On utilisera la correction de continuité suivante: $H_\alpha^* = -\frac{1}{2} + H_\alpha$. 2- FOSTER et STUART (1954) proposent d'attribuer de façon aléatoire un rang à chacune de ces valeurs, c'est ce qui est fait dans le logiciel.

PUISSANCE: Ce test est peu puissant (0.26 vs KENDALL 0.96; $N = 100$, $\alpha = 0.05$, pente = .01). Cependant son utilisation peut être d'un grand intérêt lorsque les records sont les seules données disponibles pour détecter une tendance.

RÉFÉRENCES: FOSTER, F.G. et A. STUART (1954). Distribution-free Tests in Time-Series Bases on the Breaking of Records. Journ. of the Royal Stat. Soc., Ser. B, XVI-1, 1-13.

TEST DE FOSTER ET STUART 2

FICHE # 10

OBJET DU TEST: 1- Indépendance entre les observations d'une série. 2- Tendances monotones négatives, positives (test unilatéral) et négative ou positive (test bilatéral).

CONDITIONS D'APPLICATION: Les séries contenant des saisonnalités peuvent amener des résultats particuliers pouvant biaiser les résultats du test. On n'étudiera donc aucune série à composante saisonnière à l'aide de ce test.

APPLICATION POUR LES TENDANCES: Application directe pour la détection de tendance sur la moyenne en présence de persistance dont on connaît l'amplitude.

RENSEIGNEMENTS SUPPLÉMENTAIRES: Aucune adaptation n'a été faite pour l'application en présence de saisonnalités. Dans ce cas on devra plutôt se référer à un autre type de test.

EXPRESSION ET MODE DE CALCUL: Voir FOSTER et STUART 1 fiche # 9. La statistique utilisée est encore D.

VALEURS CRITIQUES: 1-CLUIS (1986) a construit à l'aide de simulations des tables de valeur de la variance de D en présence de persistance: ρ . On notera $\sigma(D, \rho)$ les valeurs de ces tables. Les valeurs critiques sont alors $H_\alpha = t_{\alpha, n-1} \sigma(D, \rho)$.

2- Tableau sommaire des valeurs critiques pour un seuil de 95%

unilatéral								
N	10	25	50	75	100	150	200	300
ρ								
0.3	6.6	7.0	7.3	7.8	7.9	8.6	8.5	8.8
0.6	8.0	8.7	8.8	9.1	9.3	9.6	9.8	9.8
0.9	10.9	14.4	15.2	15.7	15.7	16.3	16.3	16.4
bilatéral								
N	10	25	50	75	100	150	200	300
ρ								
0.3	8.1	8.4	8.7	9.3	9.4	10.2	10.1	10.5
0.6	9.9	10.4	10.5	10.9	11.1	11.5	11.6	11.7
0.9	13.5	17.4	18.1	18.7	18.7	19.4	19.4	19.5

CORRECTIONS: 1- Aucune correction utilisée dans le logiciel. 2- Comme pour FOSTER et STUART 1, on attribuera au hasard un rang aux valeurs égales.

PUISSANCE: CLUIS (1986) a étudié la puissance de ce test pour détecter différents taux de tendances monotones. Aucun tableau ne sera présenté ici. On peut cependant citer que de façon générale, la puissance diminue avec une augmentation de la persistance et qu'elle augmente avec N. Finalement, citons le fait que pour un taux de tendance monotone de .1%, la puissance est presque nulle pour tout $N \leq 200$ et très faible pour $N = 300$.

RÉFÉRENCES: 2- CLUIS, D.A. et C. LABERGE (1986). Généralisation du test de tendance de FOSTER et STUART à des échantillons markoviens courts. INRS-Eau, rapport scientifique no. 215, 19 p.

14.6 Nombre effectif d'échantillons

En présence de persistance, les données comportent une certaine redondance. Ainsi, un échantillon persistant de taille N amènera moins d'information qu'un échantillon indépendant de même taille. On présentera ici le nombre effectif N^* , c'est-à-dire la taille d'un échantillon persistant contenant la même information qu'un échantillon indépendant de taille N, en termes de précision sur la moyenne.

Dans le cas d'un échantillon markovien on a :

$$N^* = N I$$

$$\text{où } I = \left[\frac{1 + \rho_1}{1 - \rho_1} - \frac{2}{N} \frac{\rho_1(1 - \rho_1^N)}{(1 - \rho_1)^2} \right]^{-1}$$

I s'appelle le Contenu en Information et est toujours inférieur à l'unité.

Valeurs de I

N	10	25	50	75	100	200	∞
ρ_1							
0.10	0.84	0.82	0.82	0.82	0.82	0.82	0.82
0.20	0.70	0.68	0.67	0.67	0.67	0.67	0.67
0.30	0.58	0.55	0.55	0.54	0.54	0.54	0.54
0.40	0.47	0.45	0.44	0.43	0.43	0.43	0.43
0.50	0.38	0.35	0.34	0.34	0.34	0.34	0.33
0.60	0.31	0.27	0.26	0.26	0.25	0.25	0.25
0.70	0.24	0.20	0.19	0.18	0.18	0.18	0.18
0.80	0.18	0.14	0.12	0.12	0.12	0.11	0.11
0.90	0.14	0.08	0.06	0.06	0.06	0.06	0.05
0.95	0.12	0.06	0.04	0.03	0.03	0.03	0.03

Valeurs de N^*

N	10	25	50	75	100	200
ρ_1						
0.10	8.35	20.62	41.08	61.53	81.98	163.80
0.20	6.96	16.95	33.61	50.28	66.95	133.61
0.30	5.76	13.83	27.28	40.74	54.20	108.05
0.40	4.74	11.14	21.84	32.56	43.27	86.12
0.50	3.85	8.80	17.12	25.45	33.78	67.11
0.60	3.07	6.76	12.99	19.23	25.48	50.47
0.70	2.41	4.96	9.34	13.74	18.15	35.79
0.80	1.84	3.38	6.10	8.86	11.63	22.73
0.90	1.37	2.03	3.24	4.52	5.81	11.05
0.95	1.18	1.47	2.00	2.58	3.18	5.68

14.7 Tables complémentaires des valeurs critiques

Nous présentons ici les tables des valeurs critiques qui ont été citées dans les fiches de test. On présente d'abord la table A-7 tirée de CONOVER (1971) qui pour chaque valeur de n_1 et n_2 (n et m dans la table), présente les valeurs critiques du test de MANN-WHITNEY pour les seuils (α) .001, .005, .01, .025, .05, .10. On peut alors obtenir les valeurs critiques aux seuils (α) .90, .95, .975, .99, .995, .999 en utilisant:

$$H_{\alpha} = \frac{N(N+1)}{2} - H_{1-\alpha}$$

Exemple: si $n_1 = 4$ $n_2 = 4$: la valeur critique au seuil 5% est 13 et la valeur critique au seuil 95% est donc:

$$H_{95\%} = \frac{8(9)}{2} - 13 = 23$$

La deuxième table présentée est également tirée de CONOVER (1971) (table A-12). Elle présente les valeurs critiques pour le test de KENDALL aux seuils (α) .90, .95, .975, .990 et .995. On peut obtenir les valeurs critiques pour les seuils (α) .005, .01, .025, .05 et .10 en utilisant:

$$H_{\alpha} = - H_{1-\alpha}$$

Exemple: $n = 10$, la valeur critique du seuil 95% est 19 alors qu'au seuil 5% ou a -19.

Finalement, la table A-13 de MOSTELLER et ROORKE (1973) est présentée. Elle contient les valeurs critiques pour les seuils approximatifs .025, .05, .10, .90, .95 et .975. On peut alors remarquer la symétrie par rapport à zéro des seuils α et $(1-\alpha)$.

Exemple: $n = 10$, la valeur critique pour un seuil approximatif de 95% est 0.56 alors qu'au seuil 5% elle est de -0.56.

Table 1: Valeurs critiques du test de MANN-WHITNEY pour les seuils (α) .001, .005, .01, .025, .05 et .10.

n	p	m=2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2	.001	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	.005	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
	.01	3	3	3	3	3	3	3	3	3	3	3	4	4	4	4	4	4	4	4
	.025	3	3	3	3	3	3	4	4	4	4	5	5	5	5	5	5	6	6	6
	.05	3	3	3	4	4	4	5	5	5	5	6	6	7	7	7	7	8	8	8
	.10	3	4	4	5	5	5	6	6	7	7	8	8	9	9	10	10	11	11	11
3	.001	6	6	6	6	6	6	6	6	6	6	6	6	6	6	6	7	7	7	7
	.005	6	6	6	6	6	6	7	7	7	8	8	8	8	9	9	9	9	10	10
	.01	6	6	6	6	6	7	7	8	8	8	9	9	9	10	10	11	11	11	12
	.025	6	6	6	7	7	8	8	9	10	10	11	11	12	12	13	13	14	14	15
	.05	6	7	7	8	8	9	9	10	11	11	12	12	13	14	14	15	16	16	17
	.10	7	8	8	9	10	11	12	12	13	14	15	16	17	17	18	19	20	21	22
4	.001	10	10	10	10	10	10	10	10	11	11	11	12	12	12	13	13	14	14	14
	.005	10	10	10	10	11	11	12	12	13	13	14	14	15	16	16	17	17	18	19
	.01	10	10	10	11	12	12	13	14	14	15	16	16	17	18	18	19	20	20	21
	.025	10	10	11	12	13	14	15	15	16	17	18	19	20	21	22	22	23	24	25
	.05	10	11	12	13	14	15	16	17	18	19	20	21	22	23	25	26	27	28	29
	.10	11	12	14	15	16	17	18	20	21	22	23	24	26	27	28	29	31	32	33
5	.001	15	15	15	15	15	15	16	17	17	18	18	19	19	20	21	21	22	23	23
	.005	15	15	15	16	17	17	18	19	20	21	22	23	23	24	25	26	27	28	29
	.01	15	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32
	.025	15	16	17	18	19	21	22	23	24	25	27	28	29	30	31	33	34	35	36
	.05	16	17	18	20	21	22	24	25	27	28	29	31	32	34	35	36	38	39	41
	.10	17	18	20	21	23	24	26	28	29	31	33	34	36	38	39	41	43	44	46
6	.001	21	21	21	21	21	21	23	24	25	26	26	27	28	29	30	31	32	33	34
	.005	21	21	22	23	24	25	26	27	28	29	31	32	33	34	35	37	38	39	40
	.01	21	21	23	24	25	26	28	29	30	31	33	34	35	37	38	40	41	42	44
	.025	21	23	24	25	27	28	30	32	33	35	36	38	39	41	43	44	46	47	49
	.05	22	24	25	27	29	30	32	34	36	38	39	41	43	45	47	48	50	52	54
	.10	23	25	27	29	31	33	35	37	39	41	43	45	47	49	51	53	56	58	60
7	.001	28	28	28	28	29	30	31	32	34	35	36	37	38	39	40	42	43	44	45
	.005	28	28	29	30	32	33	35	36	38	39	41	42	44	45	47	48	50	51	53
	.01	28	29	30	32	33	35	36	38	40	41	43	45	46	48	50	52	53	55	57
	.025	28	30	32	34	35	37	39	41	43	45	47	49	51	53	55	57	59	61	63
	.05	29	31	33	35	37	40	42	44	46	48	50	53	55	57	59	62	64	66	68
	.10	30	33	35	37	40	42	45	47	50	52	55	57	60	62	65	67	70	72	75
8	.001	36	36	36	37	38	39	41	42	43	45	46	48	49	51	52	54	55	57	58
	.005	36	36	38	39	41	43	44	46	48	50	52	54	55	57	59	61	63	65	67
	.01	36	37	39	41	43	44	46	48	50	52	54	56	59	61	63	65	67	69	71
	.025	37	39	41	43	45	47	50	52	54	56	59	61	63	66	68	71	73	75	78
	.05	38	40	42	45	47	50	52	55	57	60	63	65	68	70	73	76	78	81	84
	.10	39	42	44	47	50	53	56	59	61	64	67	70	73	76	79	82	85	88	91
9	.001	45	45	45	47	48	49	51	53	54	56	58	60	61	63	65	67	69	71	72
	.005	45	46	47	49	51	53	55	57	59	62	64	66	68	70	73	75	77	79	82
	.01	45	47	49	51	53	55	57	60	62	64	67	69	72	74	77	79	82	84	86
	.025	46	48	50	53	56	58	61	63	66	69	72	74	77	80	83	85	88	91	94
	.05	47	50	52	55	58	61	64	67	70	73	76	79	82	85	88	91	94	97	100
	.10	48	51	55	58	61	64	68	71	74	77	81	84	87	91	94	98	101	104	108
10	.001	55	55	56	57	59	61	62	64	66	68	70	73	75	77	79	81	83	85	88
	.005	55	56	58	60	62	65	67	69	72	74	77	80	82	85	87	90	93	95	98
	.01	55	57	59	62	64	67	69	72	75	78	80	83	86	89	92	94	97	100	103
	.025	56	59	61	64	67	70	73	76	79	82	85	89	92	95	98	101	104	108	111
	.05	57	60	63	67	70	73	76	80	83	87	90	93	97	100	104	107	111	114	118
	.10	59	62	66	69	73	77	80	84	88	92	95	99	103	107	110	114	118	122	126

Table 1 (suite)

n	p	m=2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
11	.001	66	66	67	69	71	73	75	77	79	82	84	87	89	91	94	96	99	101	104
	.005	66	67	69	72	74	77	80	83	85	88	91	94	97	100	103	106	109	112	115
	.01	66	68	71	74	76	79	82	85	89	92	95	98	101	104	108	111	114	117	120
	.025	67	70	73	76	80	83	86	90	93	97	100	104	107	111	114	118	122	125	129
	.05	68	72	75	79	83	86	90	94	98	101	105	109	113	117	121	124	128	132	136
	.10	70	74	78	82	86	90	94	98	103	107	111	115	119	124	128	132	136	140	145
12	.001	78	78	79	81	83	86	88	91	93	96	98	102	104	106	110	113	116	118	121
	.005	78	80	82	85	88	91	94	97	100	103	106	110	113	116	120	123	126	130	133
	.01	78	81	84	87	90	93	96	100	103	107	110	114	117	121	125	128	132	135	139
	.025	80	83	86	90	93	97	101	105	108	112	116	120	124	128	132	136	140	144	148
	.05	81	84	88	92	96	100	105	109	111	117	121	126	130	134	139	143	147	151	156
	.10	83	87	91	96	100	105	109	114	118	123	128	132	137	142	146	151	156	160	165
13	.001	91	91	93	95	97	100	103	106	109	112	115	118	121	124	127	130	134	137	140
	.005	91	93	95	99	102	105	109	112	116	119	123	126	130	134	137	141	145	149	152
	.01	92	94	97	101	104	108	112	115	119	123	127	131	135	139	143	147	151	155	159
	.025	93	96	100	104	108	112	116	120	125	129	133	137	142	146	151	155	159	164	168
	.05	94	98	102	107	111	116	120	125	129	134	139	143	148	153	157	162	167	172	176
	.10	96	101	105	110	115	120	125	130	135	140	145	150	155	160	166	171	176	181	186
14	.001	105	105	107	109	112	115	118	121	125	128	131	135	138	142	145	149	152	156	160
	.005	105	107	110	113	117	121	124	128	132	136	140	144	148	152	156	160	164	169	173
	.01	106	108	112	116	119	123	128	132	136	140	144	149	153	157	162	166	171	175	179
	.025	107	111	115	119	123	128	132	137	142	146	151	156	161	165	170	175	180	184	189
	.05	109	113	117	122	127	132	137	142	147	152	157	162	167	172	177	183	188	193	198
	.10	110	116	121	126	131	137	142	147	153	158	164	169	175	180	186	191	197	203	208
15	.001	120	120	122	125	128	133	135	138	142	145	149	153	157	161	164	168	172	176	180
	.005	120	123	126	129	133	137	141	145	150	154	158	163	167	172	176	181	185	190	194
	.01	121	124	128	132	136	140	145	149	154	158	163	168	172	177	182	187	191	196	201
	.025	122	126	131	135	140	145	150	155	160	165	170	175	180	185	191	196	201	206	211
	.05	124	128	133	139	144	149	154	160	165	171	176	182	187	193	198	204	209	215	221
	.10	126	131	137	143	148	154	160	166	172	178	184	189	195	201	207	213	219	225	231
16	.001	136	136	139	142	145	148	152	156	160	164	168	172	176	180	185	189	193	197	202
	.005	136	139	142	146	150	155	159	164	168	173	178	182	187	192	197	202	207	211	216
	.01	137	140	144	149	153	158	163	168	173	178	183	188	193	198	203	208	213	219	224
	.025	138	143	148	152	158	163	168	174	179	184	190	196	201	207	212	218	223	229	235
	.05	140	145	151	156	162	167	173	179	185	191	197	202	208	214	220	226	232	238	244
	.10	142	148	154	160	166	173	179	185	191	198	204	211	217	223	230	236	243	249	256
17	.001	153	154	156	159	163	167	171	175	179	183	188	192	197	201	206	211	215	220	224
	.005	153	156	160	164	169	173	178	183	188	193	198	203	208	214	219	224	229	235	240
	.01	154	158	162	167	172	177	182	187	192	198	203	209	214	220	225	231	236	242	247
	.025	156	160	165	171	176	182	188	193	199	205	211	217	223	229	235	241	247	253	259
	.05	157	163	169	174	180	187	193	199	205	211	218	224	231	237	243	250	256	263	269
	.10	160	166	172	179	185	192	199	206	212	219	226	233	239	246	253	260	267	274	281
18	.001	171	172	175	178	182	186	190	195	199	204	209	214	218	223	228	233	238	243	248
	.005	171	174	178	183	188	193	198	203	209	214	219	225	230	236	242	247	253	259	264
	.01	172	176	181	186	191	196	202	208	213	219	225	231	237	242	248	254	260	266	272
	.025	174	179	184	190	196	202	208	214	220	227	233	239	246	252	258	265	271	278	284
	.05	176	181	188	194	200	207	213	220	227	233	240	247	254	260	267	274	281	288	295
	.10	178	185	192	199	206	213	220	227	234	241	249	256	263	270	278	285	292	300	307
19	.001	190	191	194	198	202	206	211	216	220	225	231	236	241	246	251	257	262	268	273
	.005	191	194	198	203	208	213	219	224	230	236	242	248	254	260	265	272	278	284	290
	.01	192	195	200	206	211	217	223	229	235	241	247	254	260	266	273	279	285	292	298
	.025	193	198	204	210	216	223	229	236	243	249	256	263	269	276	283	290	297	304	310
	.05	195	201	208	214	221	228	235	242	249	256	263	271	278	285	292	300	307	314	321
	.10	198	205	212	219	227	234	242	249	257	264	272	280	288	295	303	311	319	326	334
20	.001	210	211	214	218	223	227	232	237	243	248	253	259	265	270	276	281	287	293	299
	.005	211	214	219	224	229	235	241	247	253	259	265	271	278	284	290	297	303	310	316
	.01	212	216	221	227	233	239	245	251	258	264	271	278	284	291	298	304	311	318	325
	.025	213	219	225	231	238	245	251	259	266	273	280	287	294	301	309	316	323	330	338
	.05	215	222	229	236	243	250	258	265	273	280	288	295	303	311	318	326	334	341	349
	.10	218	226	233	241	249	257	265	273	281	289	297	305	313	321	330	338	346	354	362

Tiré de: CONOVER, W.J. (1971). Practical Nonparametric Statistics.
2ed. John Wiley, New York, 493 p.

Table 2: Valeurs critiques du test de KENDALL pour les seuils (α)
.90, .95, .975, .990 et .995.

n	$p = .900$	$.950$	$.975$	$.990$	$.995$
4	4	4	6	6	6
5	6	6	8	8	10
6	7	9	11	11	13
7	9	11	13	15	17
8	10	14	16	18	20
9	12	16	18	22	24
10	15	19	21	25	27
11	17	21	25	29	31
12	18	24	28	34	36
13	22	26	32	38	42
14	23	31	35	41	45
15	27	33	39	47	51
16	28	36	44	50	56
17	32	40	48	56	62
18	35	43	51	61	67
19	37	47	55	65	73
20	40	50	60	70	78
21	42	54	64	76	84
22	45	59	69	81	89
23	49	63	73	87	97
24	52	66	78	92	102
25	56	70	84	98	108
26	59	75	89	105	115
27	61	79	93	111	123
28	66	84	98	116	128
29	68	88	104	124	136
30	73	93	109	129	143
31	75	97	115	135	149
32	80	102	120	142	158
33	84	106	126	150	164
34	87	111	131	155	173
35	91	115	137	163	179
36	94	120	144	170	188
37	98	126	150	176	196
38	103	131	155	183	203
39	107	137	161	191	211
40	110	142	168	198	220

Table 2 (suite)

n	p = .900	.950	.975	.990	.995
41	114	146	174	206	228
42	119	151	181	213	235
43	123	157	187	221	245
44	128	162	194	228	252
45	132	168	200	236	262
46	135	173	207	245	271
47	141	179	213	253	279
48	144	186	220	260	288
49	150	190	228	268	296
50	153	197	233	277	305
51	159	203	241	285	315
52	162	208	248	294	324
53	168	214	256	302	334
54	173	221	263	311	343
55	177	227	269	319	353
56	182	232	276	328	362
57	186	240	284	336	372
58	191	245	291	345	381
59	197	251	299	355	391
60	202	258	306	364	402

Tiré de: CONOVER, W.J. (1971). Practical Nonparametric statistics.
2ed. John Wiley, New York, 493 p.

Table 3: Valeurs critiques du test de SPEARMAN pour les seuils (α)
.025, .05, .10, .90, .95 et .975.

n	$\sum d^2$	r_s	$P(D \leq \sum d^2)$ $= P(R_s \geq r_s)$	$\sum d^2$	r_s	$P(D \geq \sum d^2)$ $= P(R_s \leq r_s)$
2	0	1	0.500	2	-1	0.500
3	0	1	0.167	8	-1	0.167
4	0	1	0.042	18	-0.8	0.167
	2	0.8	0.167	20	-1	0.042
5	0	1	0.008	34	-0.7	0.117
	2	0.9	0.042	38	-0.9	0.042
	6	0.7	0.117	40	-1	0.008
6	6	0.829	0.029	58	-0.657	0.088
	8	0.771	0.051	62	-0.771	0.051
	12	0.657	0.088	64	-0.829	0.029
7	12	0.786	0.024	88	-0.571	0.100
	18	0.679	0.055	94	-0.679	0.055
	24	0.571	0.100	100	-0.786	0.024
8	22	0.738	0.023	128	-0.524	0.098
	30	0.643	0.048	138	-0.643	0.048
	40	0.524	0.098	146	-0.738	0.023
9	38	0.633	0.025	178	-0.483	0.097
	48	0.600	0.048	192	-0.600	0.048
	62	0.483	0.097	202	-0.683	0.025
10	58	0.648	0.024	238	-0.442	0.102
	72	0.564	0.048	258	-0.564	0.048
	92	0.442	0.102	272	-0.648	0.024
11	86	0.609	0.026	312	-0.418	0.102
	104	0.527	0.050	336	-0.527	0.050
	128	0.418	0.102	354	-0.609	0.026

Tiré de: MOSTELLER, F. et R.E.K. ROURKE (1973). Study
Statistics. Addison-Wesley, Reading, 395 p.