

Frédéric BAROUSSE
Pierre GARCIA
Jean-Luc LOUBET

Département Informatique et Gestion
Projet 3^{ème} Année

ORSTOM
**Logiciel d'ajustements
statistiques**
Rapport de Synthèse

Tuteur : Mme Christine LAVIT

Février 1992

SOMMAIRE

REMERCIEMENTS.....	2
INTRODUCTION.....	3
L'ORSTOM.....	4
NOTIONS DE STATISTIQUES.....	5
LE PROBLEME.....	8
A-Le projet	
1-Le point de départ de l'étude.....	8
2-Les objectifs de l'étude.....	8
3-Champ de l'étude.....	9
4-Modalités et moyens.....	9
B-La démarche	
1-Etude de l'environnement.....	10
2-Etude du rapport d'évaluation.....	10
3-Les spécifications détaillées.....	10
4-Rédaction des documents.....	10
LA SOLUTION.....	11
A-Les données.....	11
1-Deux catégories d'informations.....	11
2-Le modèle conceptuel des données.....	12
B-Les traitements.....	13
1-Le modèle conceptuel des traitements.....	13
2-Les modèles fonctionnels des traitements.....	14
3-Description des principales fonctionnalités.....	25
C-Critique de la solution.....	27
D-Chiffrage de la solution.....	28
E-Avenir du projet.....	29
CONCLUSION.....	30
ECHEANCIER.....	31
BIBLIOGRAPHIE.....	32

REMERCIEMENTS

Nous voulons, avant tout, remercier Hélène LUBES, ingénieur hydrologue à l'ORSTOM Montpellier de nous avoir proposé ce projet et d'avoir été une interlocutrice disponible, efficace, qui a tout mis en oeuvre pour faciliter notre tâche.

Nous tenons à témoigner notre reconnaissance à Patrick RAOUS et François DELCLAUX pour nous avoir fait profiter de leurs connaissances en informatique et statistique.

Enfin, nous remercions Christine LAVIT, notre professeur et tuteur de projet, ainsi que l'ensemble des enseignants du département, pour nous avoir conseillés et guidés tout au long de notre travail.

INTRODUCTION

Le projet sur lequel nous avons travaillé, en collaboration avec le laboratoire d'hydrologie de l'ORSTOM de Montpellier, s'intègre dans la mission d'appui informatique à la réalisation des programmes de recherche, pour la plupart développés à l'étranger.

Le laboratoire d'hydrologie de l'ORSTOM et le laboratoire d'hydrologie et de modélisation (LHM) de l'université de Montpellier II ont réalisé une étude comparative d'évaluation de quatre logiciels d'ajustements de lois statistiques à des variables hydrologiques sur micro ordinateur compatible PC.

Cette étude montre l'inadéquation entre les différents logiciels étudiés et les besoins en outils statistiques des hydrologues de l'ORSTOM.

Les avis apportés dans ce rapport débouchent sur des propositions pour la conception d'un nouveau logiciel dont la particularité sera de guider l'utilisateur tout au long de sa démarche statistique.

Notre travail consistait à réaliser l'étude préalable et à concevoir les spécifications détaillées du futur produit.

L'ORSTOM

L'ORSTOM et les pays du tiers monde

L'ORSTOM, rebaptisé Institut Français de Recherche Scientifique pour le Développement en Coopération, a pour mission de promouvoir la recherche pour le développement des pays du tiers monde. Il possède plus de cinquante antennes outre-mer et intervient dans plusieurs pays du globe.

L'institut développe, conjointement avec ses partenaires, des programmes de recherche à long terme. Des actions sont menées dans les principaux domaines scientifiques et technologiques concernant les activités de ces pays.

L'ORSTOM à Montpellier

Le centre ORSTOM de Montpellier intervient sur six pôles de recherche : la santé, l'hydrologie, la biotechnologie, les ressources génétiques, l'océanographie et la phytopathologie.

Les travaux de recherche et de développement sont financés par des subventions de l'état et par des conventions passées avec des entreprises, des pays étrangers et des organismes internationaux (ONU, FAO, UNESCO, ...)

Le laboratoire d'hydrologie

Les recherches en hydrologie sont menées dans les secteurs suivants : les relations Continent-Atmosphère, la géodynamique de l'atmosphère continentale, le processus de transformation Sol-Eau-Plante-Atmosphère, l'évaluation et la gestion des ressources en eau.

Le laboratoire est chargé d'apporter un appui logistique, informatique et documentaire à la réalisation de ces programmes de recherche, pour la plupart développés à l'étranger.

NOTIONS DE STATISTIQUES

L'étude des risques associés à un événement hydrologique (crue, sécheresse...) passe toujours par l'analyse statistique d'observations. L'utilisation de formules empiriques régionales ou d'abaques n'échappe pas à cette règle : pour les établir, il a fallu procéder à l'analyse statistique des résultats antérieurs.

Rappelons que l'analyse statistique des observations du passé a pour objectif d'obtenir des informations sur la population d'où elles sont tirées, afin de pouvoir énoncer des probabilités concernant l'avenir. La démarche statistique, qui suppose la stabilité de la population, s'effectue en trois étapes :

-Sélection d'observations et de critères qui décrivent le phénomène étudié. Pour chaque critère l'échantillon ainsi constitué doit respecter certaines propriétés afin qu'il soit possible d'en tirer des informations concernant la population.

-Ajustement d'une loi de probabilité théorique à la distribution de fréquences de cet échantillon. Il existe un éventail très large de lois théoriques et de méthodes d'ajustement de ces lois.

-Le choix d'une loi de probabilité pour décrire le phénomène étudié permet d'énoncer des probabilités concernant l'avenir.

L'hydrologue emploie des concepts spécifiques, le statisticien a son propre langage. Il semble donc important de mettre au point certaines notions de base nécessaires à la compréhension du projet.

Qualité de l'échantillon

Pour être représentatif de la population dont il est tiré, l'échantillon des observations doit présenter un certain nombre de caractéristiques bien précises.

Les observations doivent être :

-Aléatoires : C'est à dire, en hydrologie, être le résultat de fluctuations naturelles.

-Indépendantes : La valeur d'une observation ne doit pas être influencée par la valeur de l'observation précédente au sens chronologique.

-Homogènes : C'est à dire provenant de la même population.

-Stationnaires : Les variables appartenant à une série chronologique sont dites stationnaires lorsque leurs caractéristiques statistiques (moyenne, autocovariance...) ne changent pas avec les saisons. Les variations naturelles dues aux saisons sont neutralisées en découpant l'année en périodes pendant lesquelles on considère que la variable est stationnaire.

L'ajustement

L'ajustement a pour but de déterminer, à partir des données de l'échantillon, les paramètres d'une loi au moyen de méthodes d'estimation.

Les méthodes d'estimation des paramètres de lois

Les lois statistiques possèdent un ou plusieurs paramètres qui, dans la population d'où est tiré l'échantillon ont des valeurs données mais inconnues. Il s'agit d'estimer ces valeurs à partir des observations de l'échantillon. Les estimations réalisées aboutissent à des résultats différents d'un échantillon à l'autre, c'est la raison pour laquelle, à partir d'un échantillon, on donne non seulement une valeur du paramètre mais aussi un intervalle de confiance qui a une forte probabilité de contenir la vraie valeur du paramètre.

Les quantiles

On appelle quantile x_p de la loi de probabilité théorique, la valeur numérique prise par la variable aléatoire X pour une probabilité de non dépassement donnée p . C'est à dire que la probabilité que X prenne une valeur inférieure ou égale à x_p est égale à p . x_p se calcule grâce à la loi théorique de la variable aléatoire X .

La formule de fréquence empirique ou "plotting-position"

Lorsque les valeurs de l'échantillon sont classées par ordre croissant (ou décroissant), à une valeur x_i de l'échantillon on fait correspondre par une formule dite de "plotting-position" une fréquence empirique de non dépassement (cette valeur dépend du nombre de valeurs inférieures ou égales à x_i et de la taille de l'échantillon). Ces quantiles empiriques sont par la suite comparés aux quantiles théoriques déduits de l'ajustement.

Les tests d'ajustement

A chaque ajustement (une loi, une méthode) correspond un jeu de paramètres unique. Il reste alors à choisir parmi les lois proposées celle qui paraît la mieux adaptée à l'échantillon traité. Ce choix peut reposer sur un test statistique ou sur une évaluation visuelle de la pertinence de l'ajustement lorsqu'on superpose la loi ajustée aux données et les fréquences empiriques de non dépassement (quantiles empiriques et théoriques).

LE PROBLEME

A-LE PROJET

Notre travail consiste à réaliser l'étude préalable et détaillée d'un logiciel complet d'ajustement de lois statistiques sur des variables hydrologiques à partir du rapport d'évaluation des quatre logiciels d'ajustements testés par l'ORSTOM et le LHM.

1-Le point de départ

Les logiciels d'ajustements testés se sont avérés à l'usage, lourds et limités. Un manque d'ergonomie, de convivialité, le besoin d'une plus grande diversité de lois et surtout la nécessité de fournir des graphiques de grande qualité, ont remis en cause leur utilisation pour des ajustements d'échantillons hydrologiques.

Les hydrologues ont jugé nécessaire de reconsidérer les fonctionnalités de ces logiciels. Il a été décidé de réaliser un logiciel complet dont les qualités premières seraient d'une part, d'offrir une gestion transparente des échantillons et des résultats des ajustements et, d'autre part, de guider l'utilisateur tout au long de sa démarche statistique.

2-Les objectifs de l'étude

- Etre à l'écoute des hydrologues afin de leur proposer une solution adéquate.
- Réaliser l'étude préalable et détaillée du futur logiciel afin de permettre son développement.
- Estimer les charges en matériel et en personnel pour effectuer ce développement.
- Donner un avis sur une éventuelle portabilité sur station de travail.
- Proposer un produit évolutif, c'est dire permettant l'ajout aisé de nouvelles lois et méthodes de calcul.

3-Champs de l'étude

Le travail qui nous a été demandé nous a amené à aborder les thèmes suivants :

- Etude des principales notions de statistiques et d'hydrologie nécessaires à la compréhension du problème.
- Prise en compte des nouvelles requêtes des utilisateurs.
- Conception et analyse détaillée d'une solution adaptée à l'attente des hydrologues.
- Etude de la portabilité de PC sur les stations de travail.
- Evaluation des logiciels d'aide au développement.
- Préparation du futur développement.
- Rédaction du rapport technique présentant notre travail et qui constitue un support pour le développement du logiciel.

4-Les modalités et les moyens

Hélène LUBES, ingénieur hydrologue à l'ORSTOM Montpellier a été notre interlocuteur principal. Elle sera l'un des futurs utilisateurs de notre produit.

François DELCLAUX, responsable de l'unité informatique du laboratoire nous a conseillé sur tous les problèmes matériels que nous avons rencontrés.

Patrick RAOUS, ingénieur de recherche en hydrologie, nous a fait profiter de ses connaissances tant sur le plan statistique que sur le plan informatique.

B-LA DEMARCHE

1-Etude de l'environnement

Notre premier contact avec le projet a eu lieu lors d'un entretien avec Hélène LUBES qui nous a présenté le laboratoire d'hydrologie de Montpellier, ainsi que les quatre logiciels d'ajustement disponibles à l'ORSTOM.

Elle nous a expliqué les notions d'hydrologie et de statistiques nécessaires à la compréhension du problème, et nous a fourni toutes les informations relatives aux quatre logiciels d'ajustement ainsi que le rapport de synthèse de l'étude comparative réalisé sur ces quatre logiciels par l'ORSTOM et le LHM.

2-Etude du rapport d'évaluation

Nous avons lu le rapport d'évaluation, tout en essayant de comprendre les concepts fondamentaux auxquels il fait référence. Nous ne cachons pas que nous ne les avons pas approfondis et que nous n'en maîtrisons pas toutes les subtilités. Cependant, nous avons cerné l'objectif de notre travail et défini la logique de l'étude à laquelle nous devons nous livrer.

L'étude de ce rapport, ainsi que les différents entretiens que nous avons eu avec les utilisateurs nous a permis de définir de façon détaillée l'architecture, le fonctionnement et les différents traitements du logiciel répondant le plus aux besoins des hydrologues de l'ORSTOM.

3-Les spécifications détaillées

Le modèle logique défini, nous avons décrit chacun des traitements de chacune des procédures à l'aide de fiches précisant : les événements qui déclenchent ces traitements, les procédures auxquelles ils appartiennent, les données qui sont affichées ou saisies à l'écran, les sous-tâches qui les composent et les résultats.

4-Rédaction des documents

Les documents du rapport technique ont été rédigés afin de permettre une programmation immédiate. Nous avons veillé à ce qu'ils soient à la fois clairs et précis. Ils permettent d'inclure d'éventuelles modifications avec une relative facilité.

LA SOLUTION

A-LES DONNEES

1-Deux catégories d'informations

Par la nature même du système étudié, deux catégories d'informations doivent être distinguées :

-Les informations hydrologiques et statistiques. Ce sont les données physiques recueillies par l'hydrologue.

-Les informations liées aux traitements du logiciel. Ce sont les paramètres des lois et les différentes variables générées par les traitements.

Ces dernières constituent les informations participant au fonctionnement des lois ou résultant de l'application de ces modèles aux données de la première catégorie.

a-Les informations hydrologiques et statistiques

-Echantillon : Cette information contient l'identification et les caractéristiques d'un échantillon.

-Données : Ce sont les valeurs des mesures recueillies pour un échantillon.

b-Les informations liées aux traitements

-Session : Cette information contient les résultats d'un ajustement.

-Paramètres : A chaque loi est associé un jeu de paramètres. Ils sont calculés à l'aide des méthodes d'ajustements.

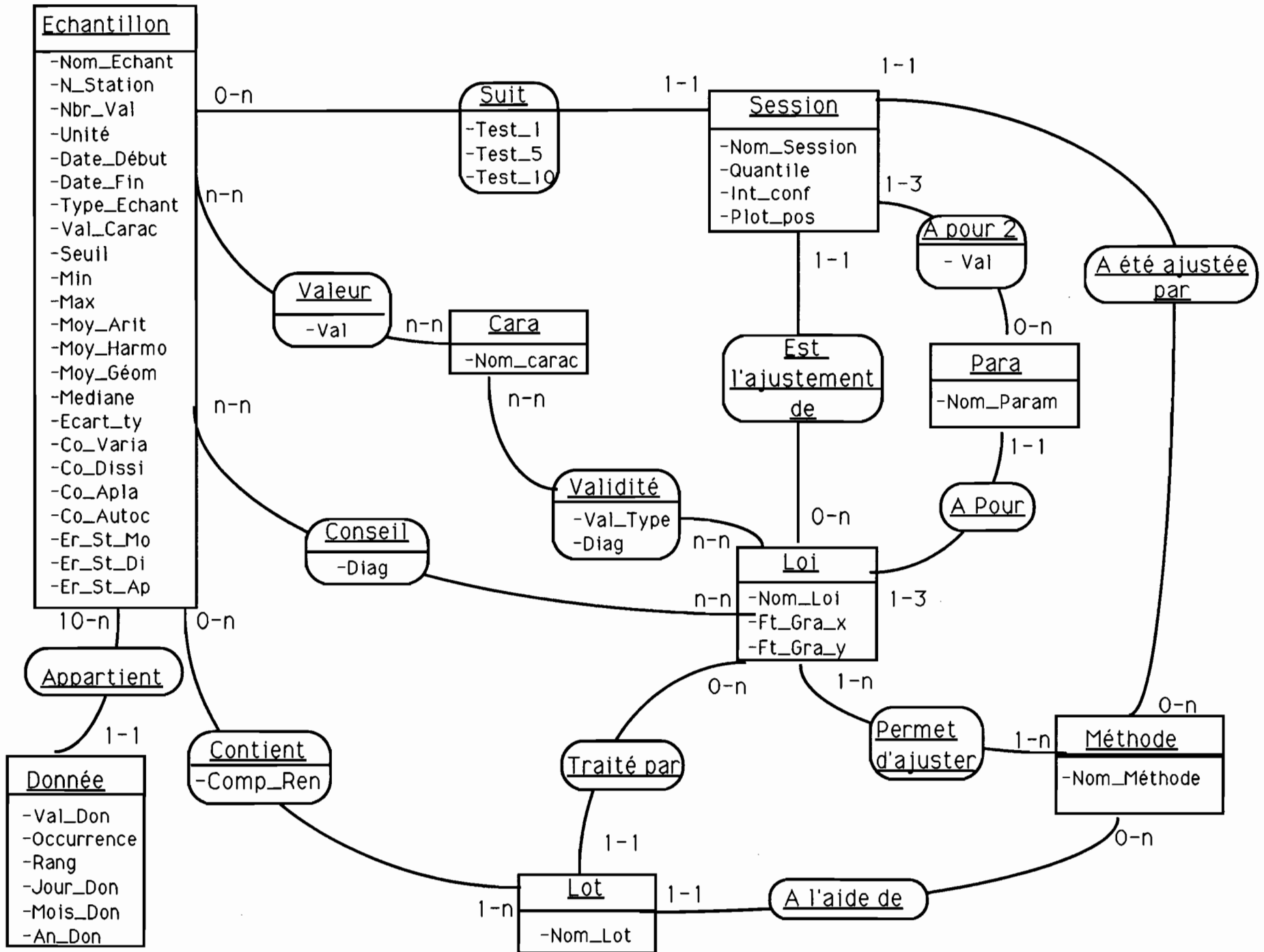
-Loi : Cette information fait référence au programme d'ajustement utilisé par une session.

-Méthode : Pour chaque loi, il existe une ou plusieurs méthodes d'évaluation des paramètres.

-Caract : Les caractéristiques sont des données calculées à partir des valeurs de l'échantillon. Elles permettent de définir les conseils guidant l'utilisateur dans le choix des lois à ajuster avec cet échantillon.

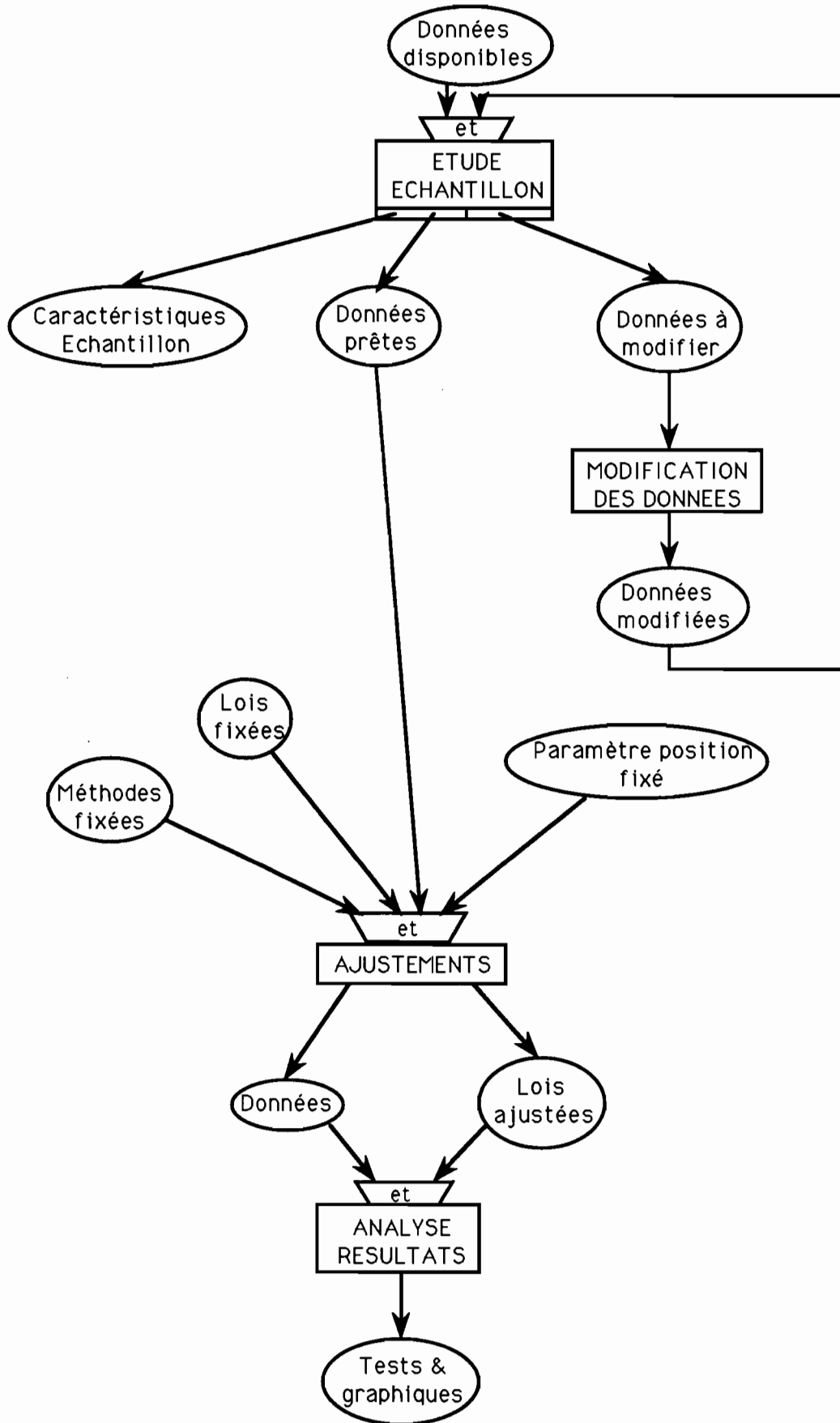
-Lot : Lors d'un traitement par lot, l'utilisateur ajuste plusieurs échantillons suivant une même loi à l'aide d'une même méthode.

2-Le Modèle Conceptuel des Données



B-LES TRAITEMENTS

1-Le Modèle Conceptuel des Traitements

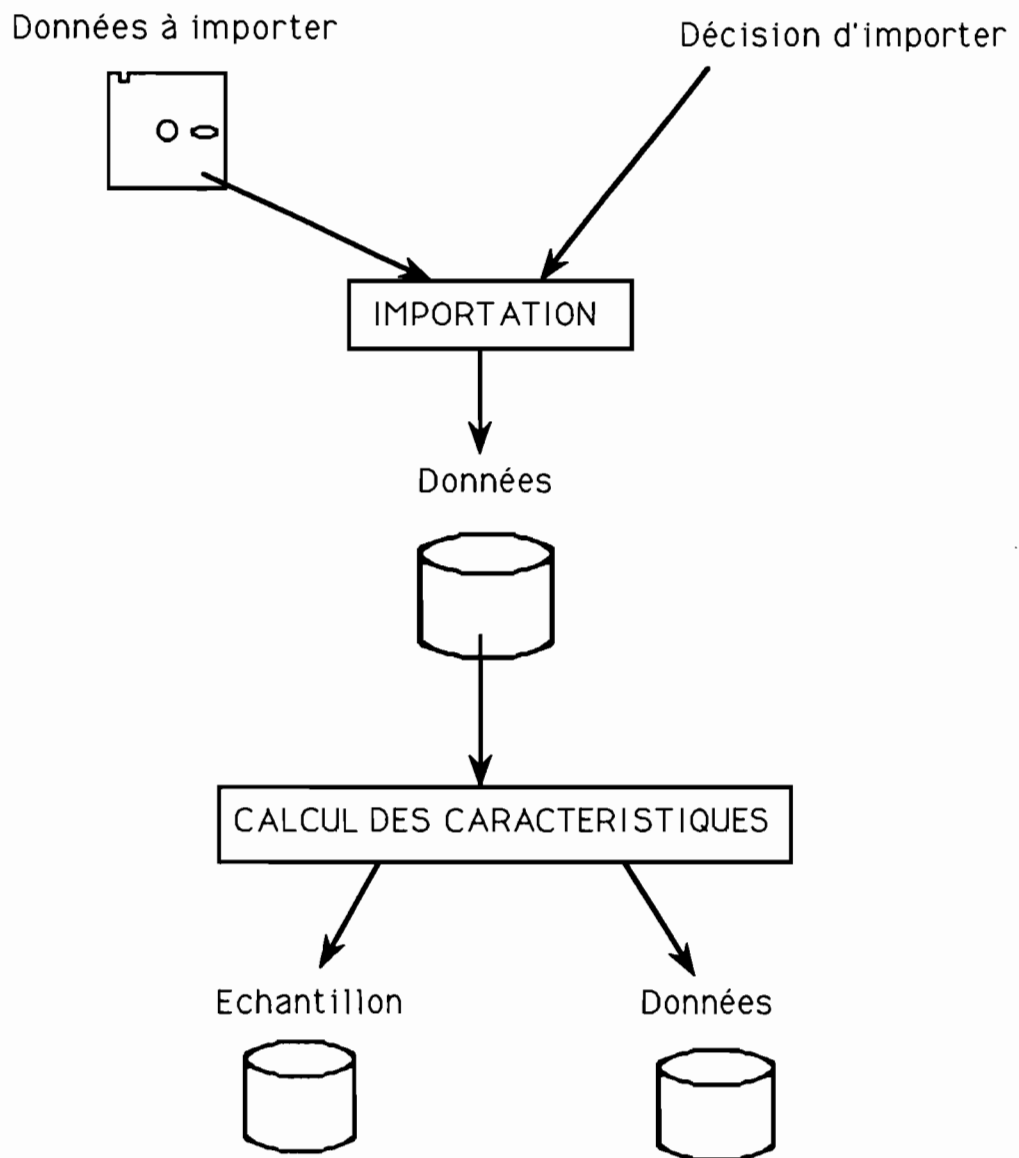


2-Les Modèles Fonctionnels des Traitements

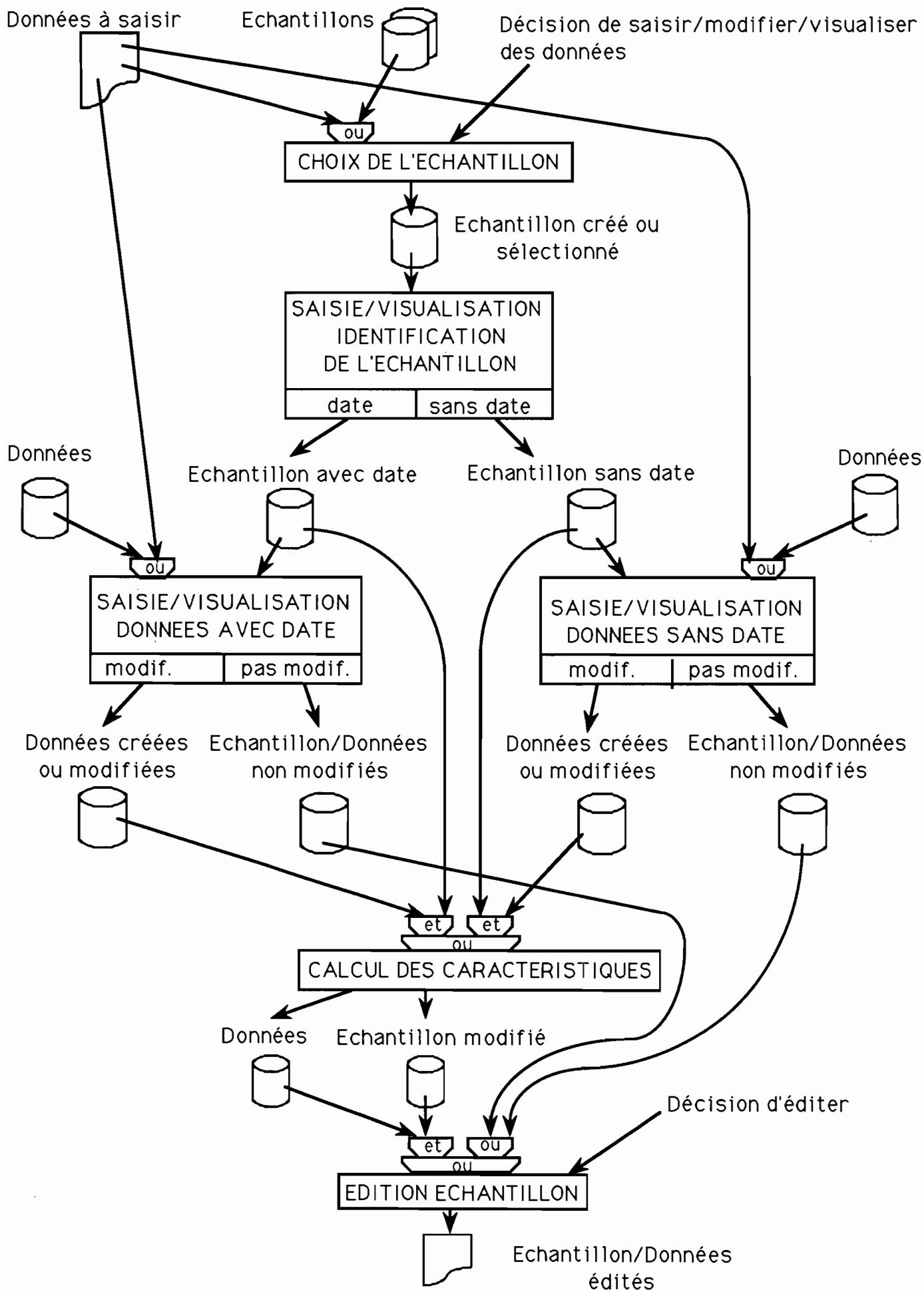
Nous avons choisi, afin de ne pas alourdir ce rapport, de présenter uniquement quelques exemples de traitements du modèle fonctionnel :

- importation de fichiers.
- saisie et visualisation des données.
- étude des caractéristiques de l'échantillon.
- histogramme.
- ajustement de lois.
- traitement par lot.
- consultation de sessions.

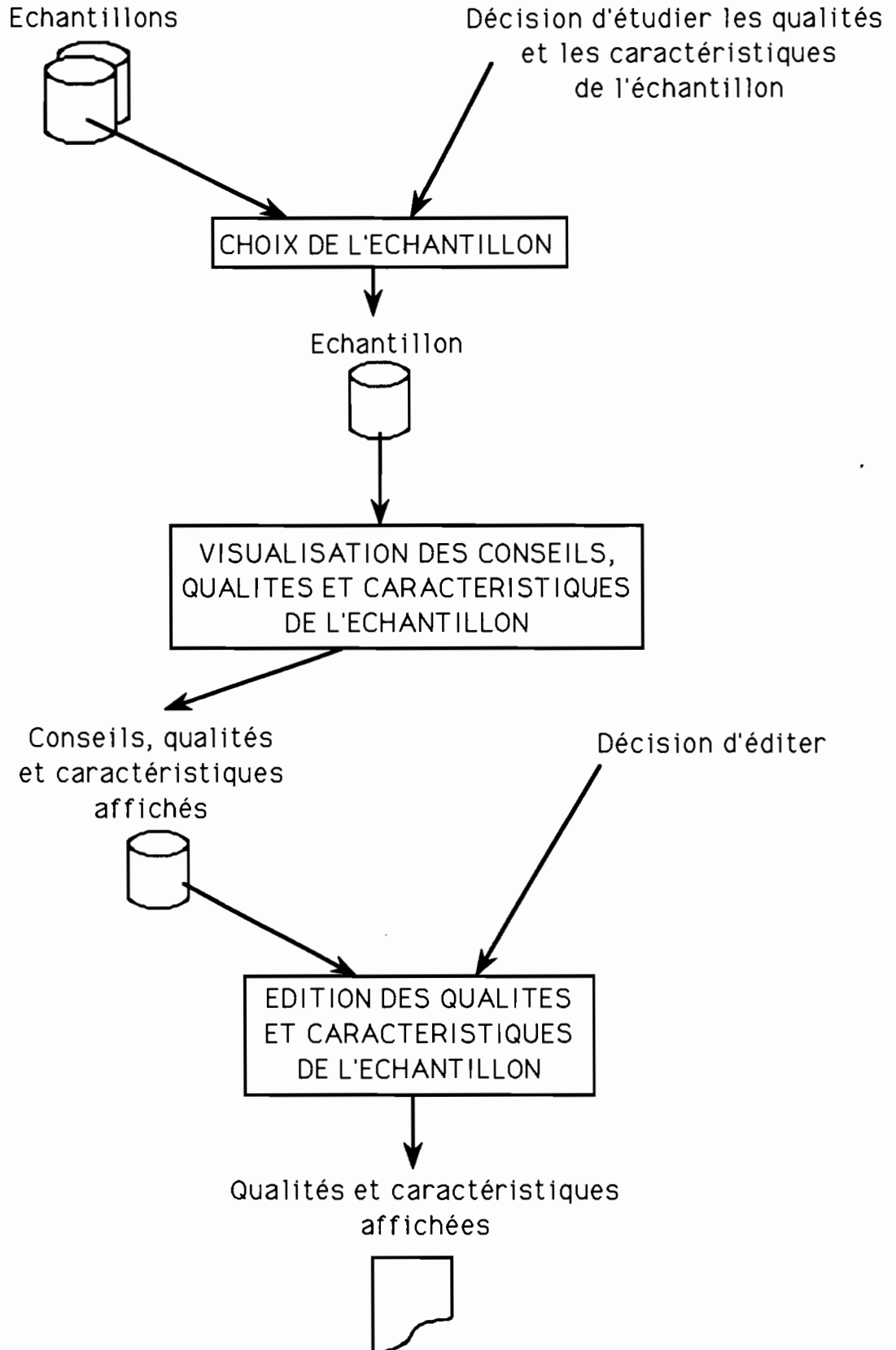
IMPORTATION



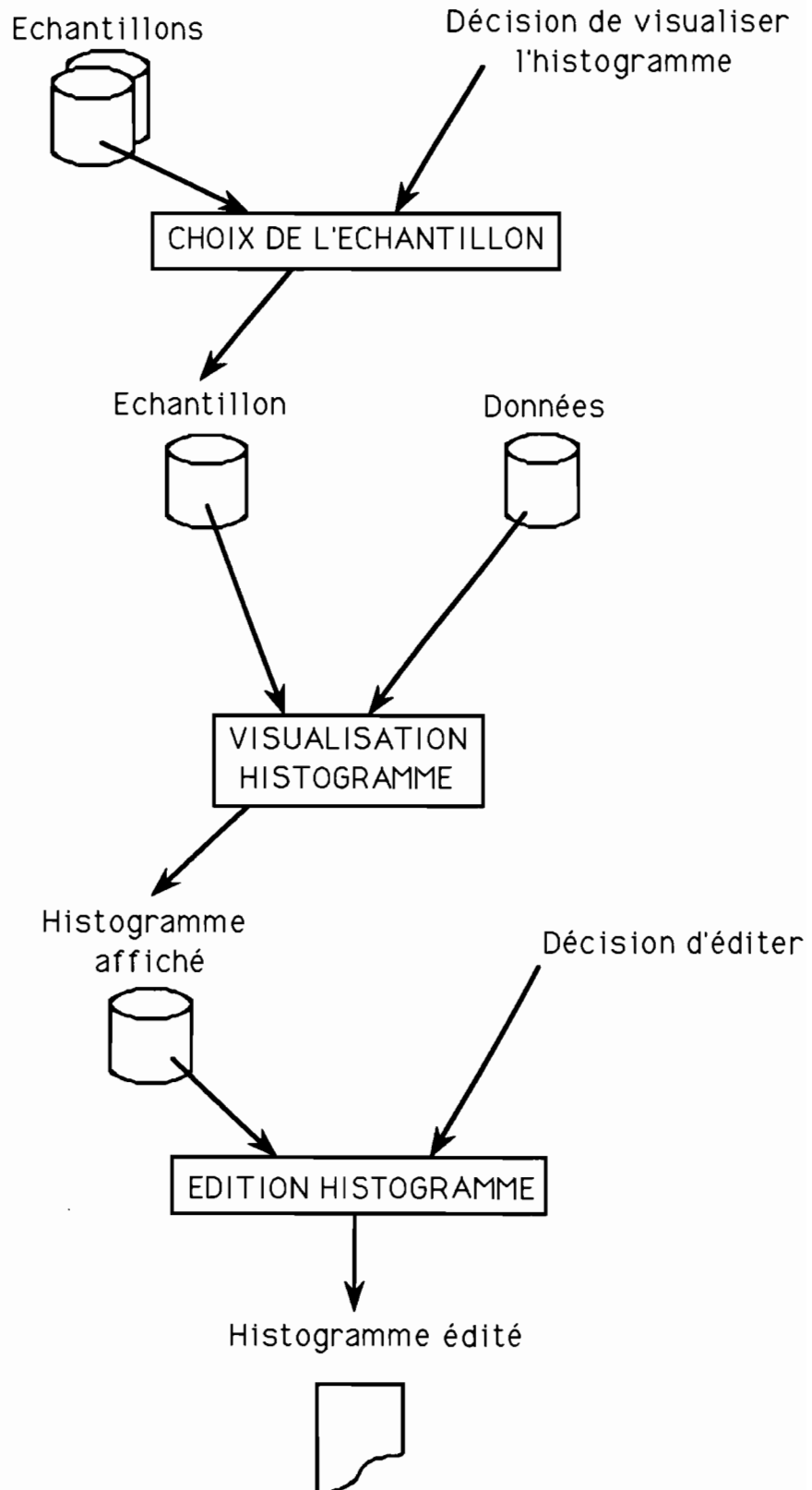
SAISIE/VISUALISATION DES DONNEES



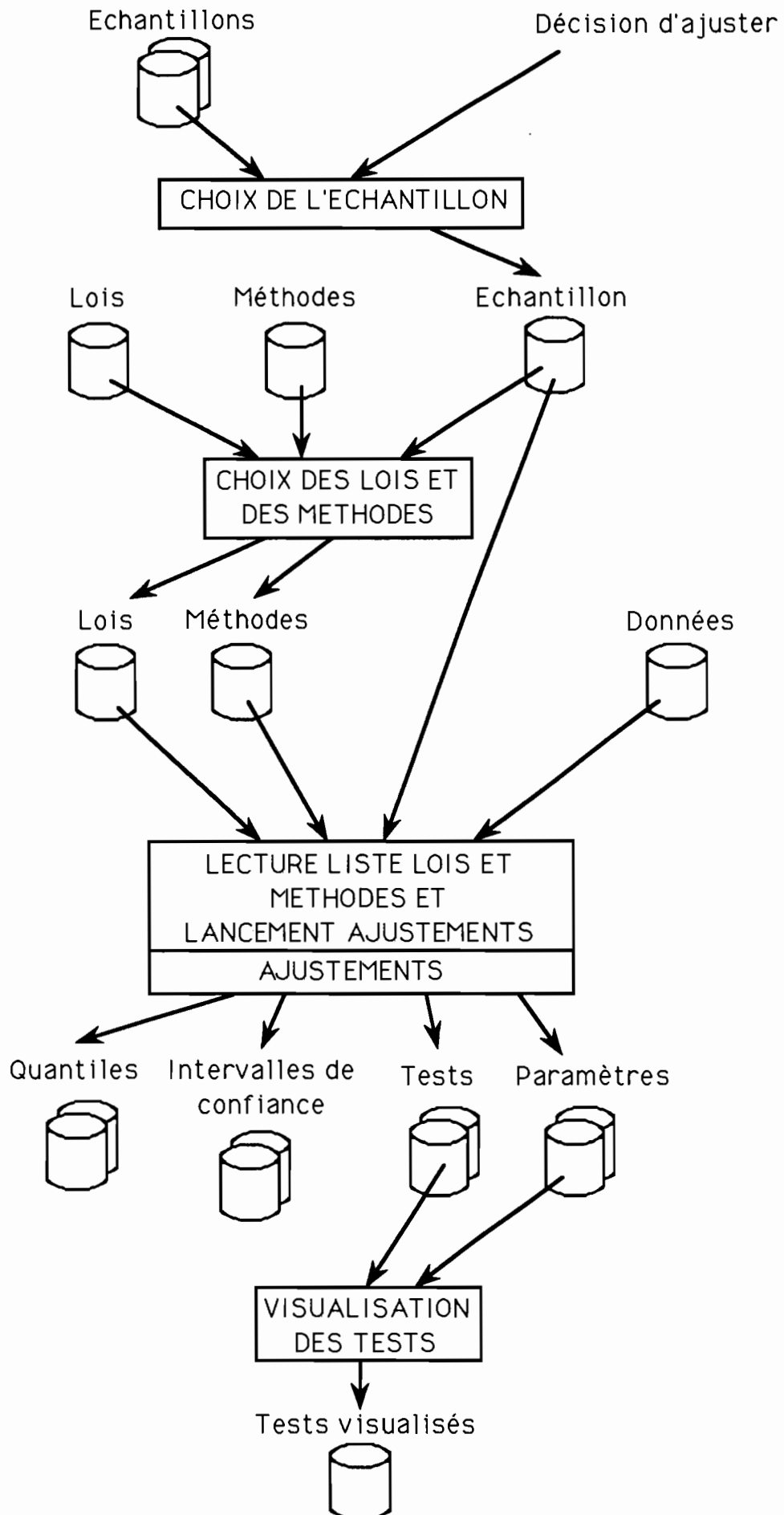
ETUDE DES QUALITES ET CARACTERISTIQUES DE L'ECHANTILLON

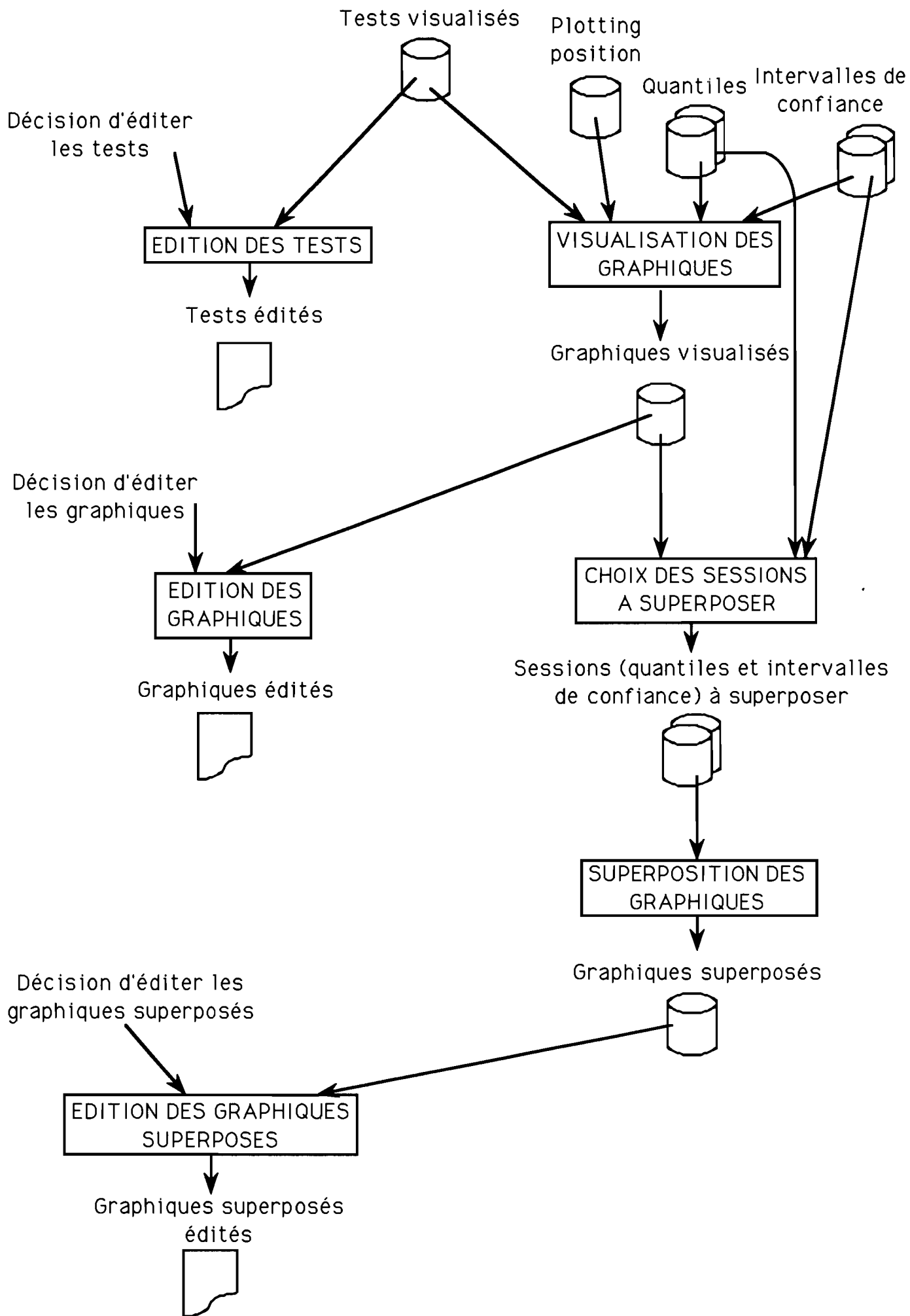


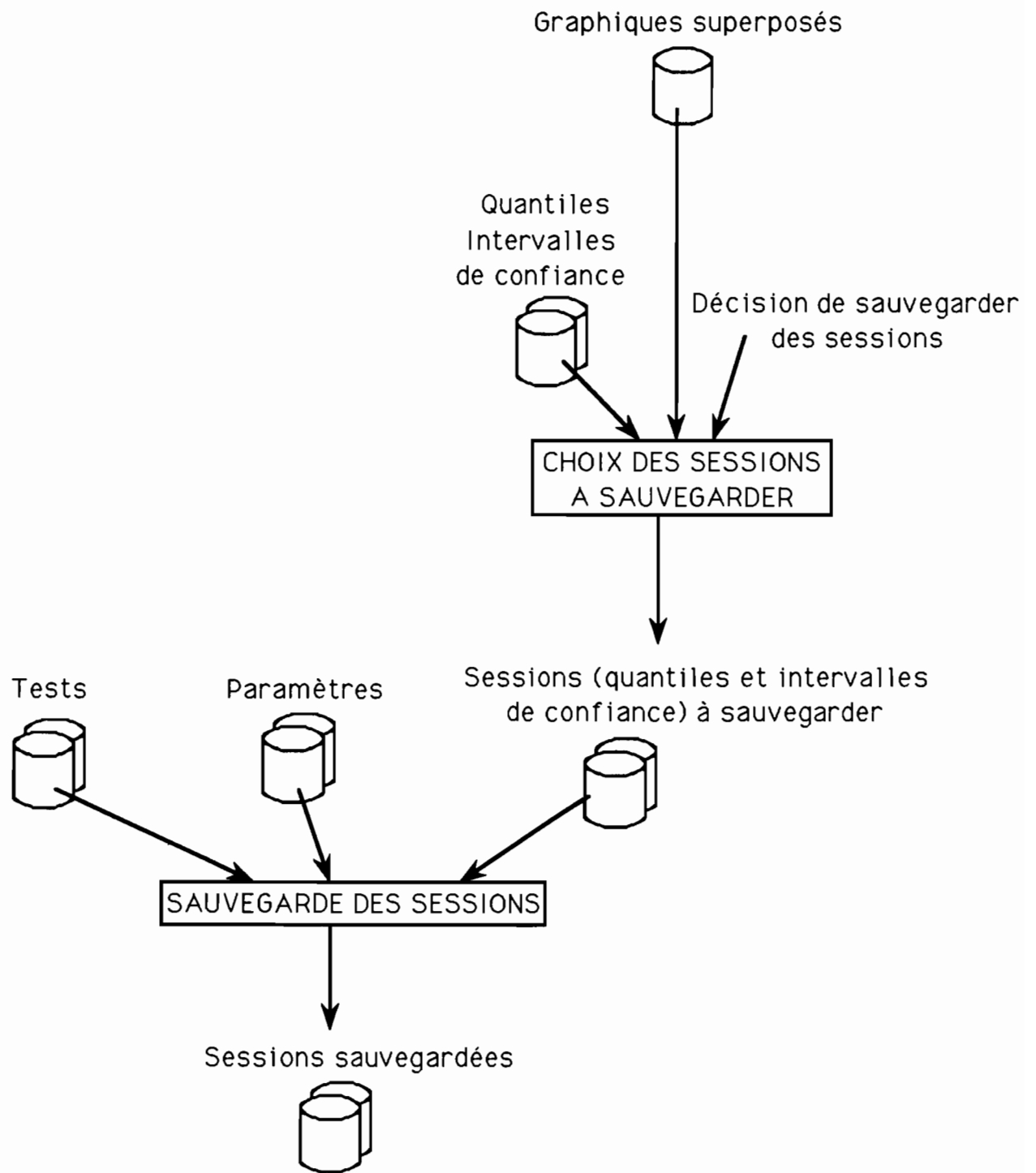
HISTOGRAMME



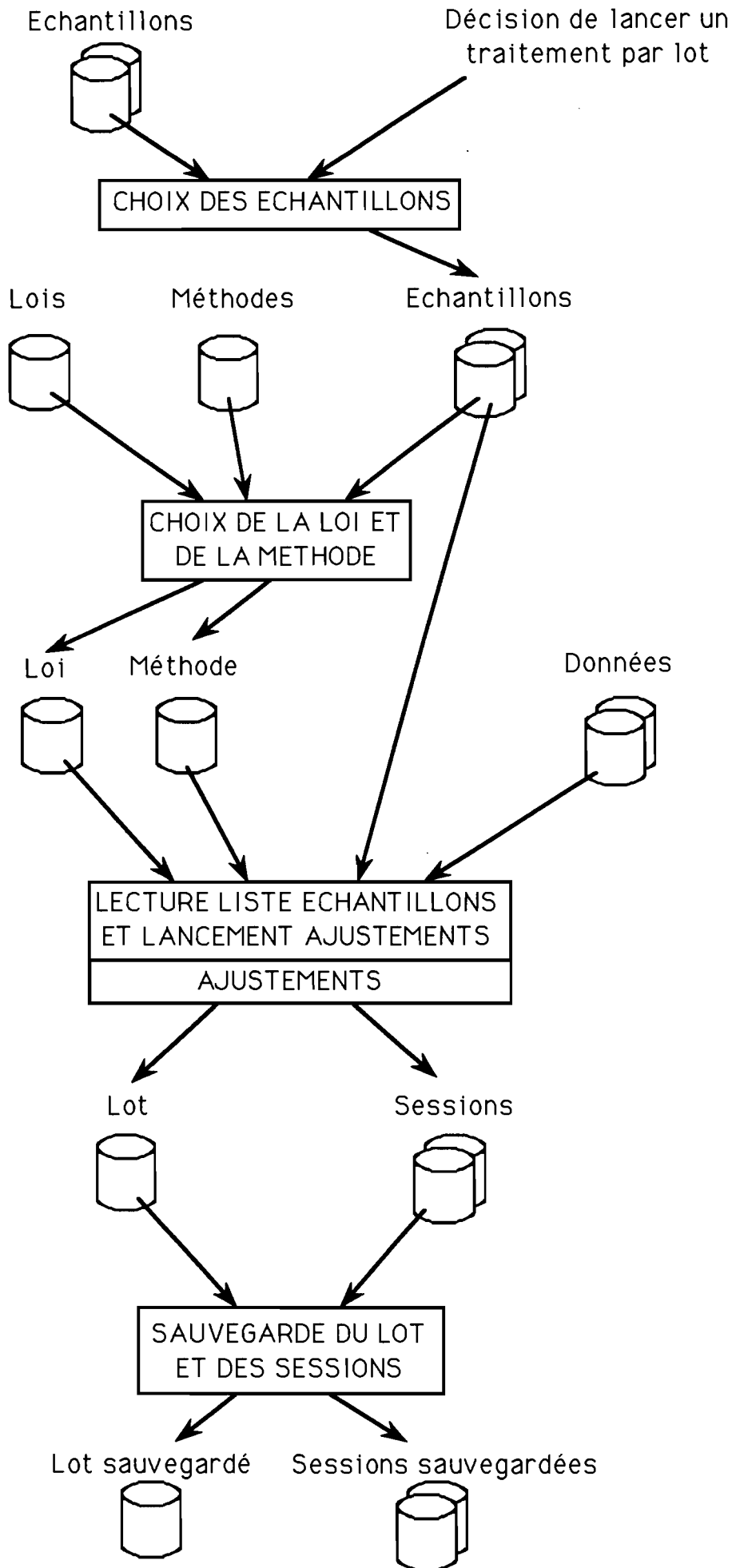
AJUSTEMENTS DE LOIS



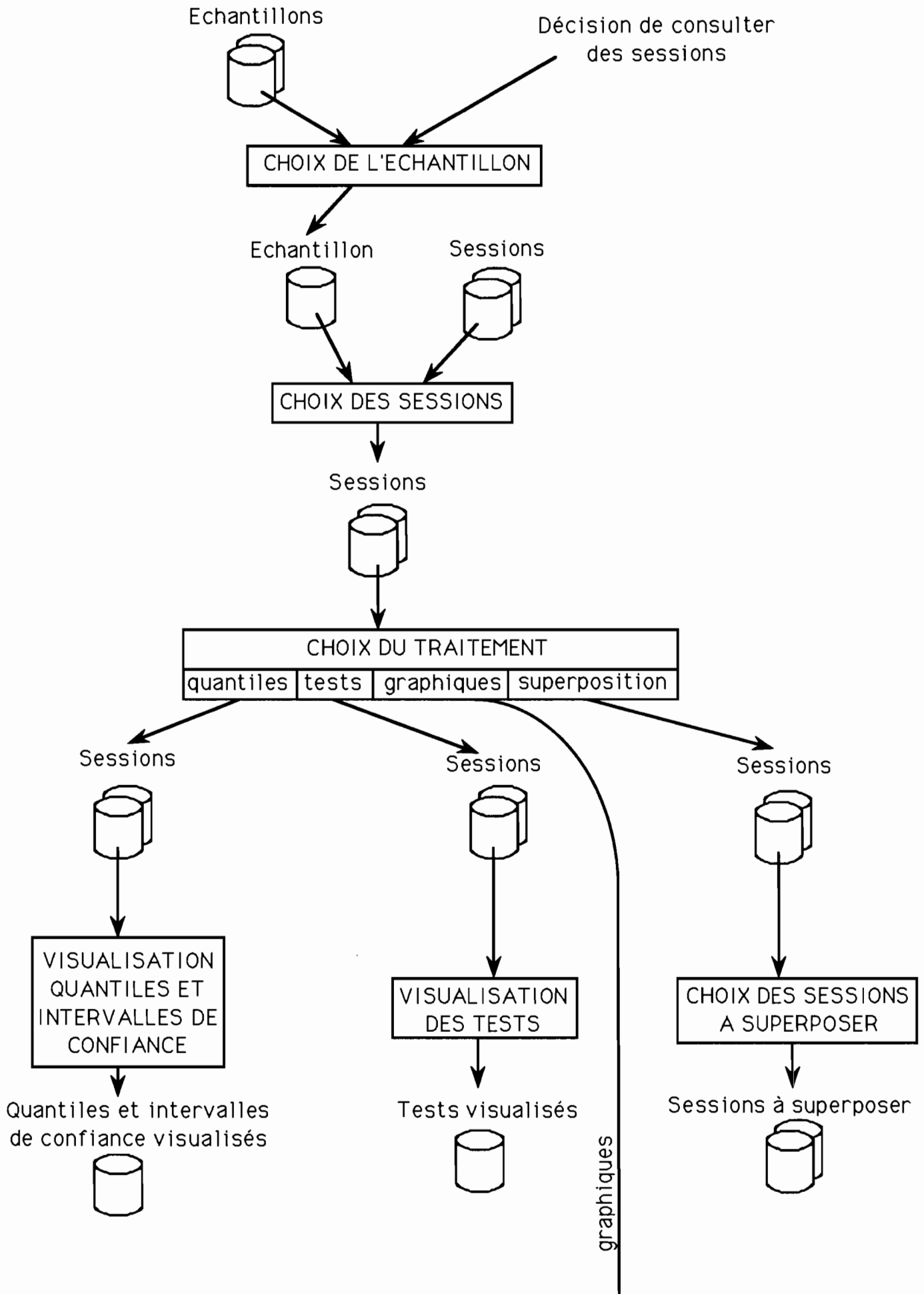


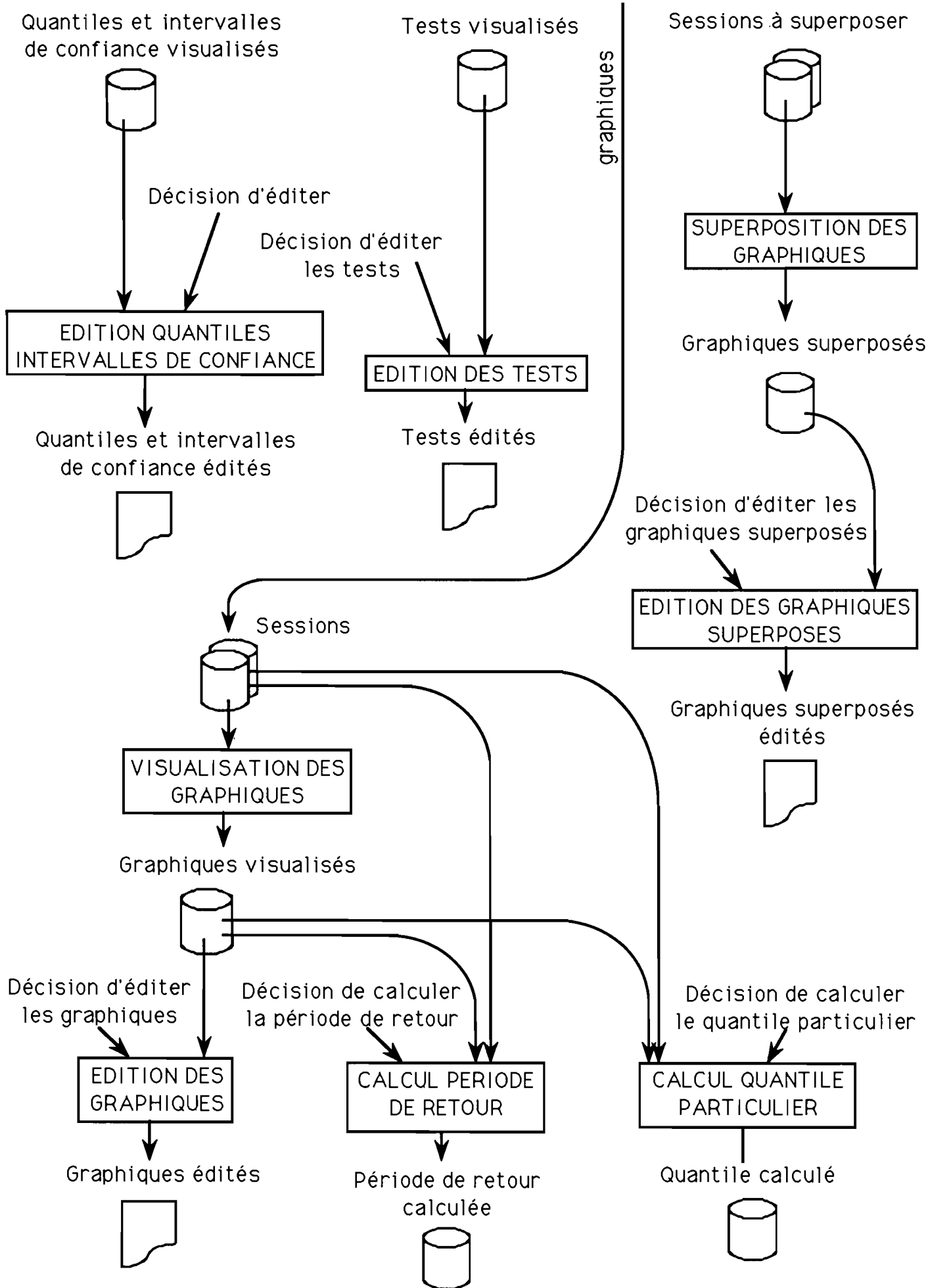


TRAITEMENT PAR LOT



CONSULTATION DES SESSIONS





3-Description des principales fonctionnalités

Gestion des échantillons

Cette fonctionnalité permet la préparation des échantillons de données qui serviront par la suite à la réalisation des différents ajustements.

Etude des caractéristiques

Cette fonctionnalité permet de visualiser les caractéristiques et les qualités d'un échantillon de données (écart-type, valeurs extrêmes, variance, etc...). Elle permet donc à l'utilisateur de déceler d'éventuelles valeurs aberrantes mais aussi d'orienter ses choix pour un éventuel ajustement.

Ajustement de lois

Ce traitement permet de réaliser les ajustements pour un échantillon de données et de consulter et comparer les résultats de ces ajustements au moyen de paramètres, de tests d'adéquation et de graphiques. Les ajustements alors réalisés pourront être sauvegardés sous forme de sessions. 17 lois distinctes et 4 méthodes sont offertes (cf rapport technique page 20).

Traitement par lot

Le traitement par lot permet de réaliser le même ajustement (même loi, même méthode), sur divers échantillons afin de déceler d'éventuelles similarités. Les différentes sessions réalisées sont automatiquement sauvegardées et pourront être visualisées par l'exploitation des résultats.

Exploitation des résultats

Cette fonctionnalité permet de consulter les différentes sessions réalisées préalablement (traitement par lot ou ajustement de lois). Les résultats des sessions sont visualisables sous forme de graphiques, de quantiles, d'intervalles de confiance, de paramètres et de tests d'adéquation. Elle permet de plus d'éditer un dossier complet du travail réalisé sur un échantillon.

C-CRITIQUE DE LA SOLUTION

La solution proposée a consisté à reprendre les différentes fonctionnalités des logiciels d'ajustements de l'étude d'évaluation, de concevoir une interface utilisateur et d'insérer des lois supplémentaires permettant d'améliorer et faciliter le travail statistique.

Cette solution a pour avantage :

-la souplesse dans l'acquisition des données qui peuvent avoir diverses origines :

- Importation de différents types de fichiers.
- Saisie à l'aide de grilles.

-l'utilisateur n'a pas recours à un éditeur de texte pour la création et la consultation de la base de donnée.

-la possibilité de modifier à tout moment et de façon simple les données de la base.

-la gestion transparente de tous les fichiers.

-la diversité des lois et des méthodes, qui permet d'étudier de façon significative une grande gamme d'échantillons hydrologiques.

-de fournir une maquette qui permet d'évaluer le futur logiciel et d'offrir un point de départ au développement (écrans réalisés avec High Screen).

D-CHIFFRAGE DE LA SOLUTION

Les charges de travail pour la réalisation du logiciel sont estimées à 9 mois/Homme de programmation sous les conditions suivantes :

-Les programmes seront écrits en langage PASCAL. On a choisi la version 5 du TURBO PASCAL qui est actuellement disponible au laboratoire d'hydrologie de l'ORSTOM.

-Les écrans et leurs enchaînements seront réalisés à l'aide du logiciel HIGH SCREEN version 5 également disponible à l'ORSTOM.

-Les graphiques et leurs éditions seront réalisés à l'aide d'un ou plusieurs utilitaires de développement dont le choix n'est pas encore arrêté (cf rapport technique page 54).

-Le logiciel sera réalisé et installé dans un premier temps sur un PS avec écran graphique déjà disponible au laboratoire.

E-AVENIR DU PROJET

1-A court terme

La réalisation proprement dite du logiciel d'ajustement se décomposera en deux parties :

-le développement des ajustements qui sera réalisé par les chercheurs de l'ORSTOM.

-le développement de l'interface utilisateur, c'est à dire les programmes de gestion des données et exploitation des résultats des ajustements. Ce développement donnera lieu à un projet de deuxième année et sera donc réalisé par trois étudiants du département informatique et gestion de l'ISIM.

2-A long terme

Le développement terminé, le logiciel sera dans un premier temps utilisé par les chercheurs du laboratoire d'hydrologie de l'ORSTOM de Montpellier puis diffusé à l'ensemble des chercheurs de l'ORSTOM.

Les difficultés de transfert des fichiers et notamment des graphiques ne permet pas d'envisager une portabilité de notre logiciel de PC vers les stations de travail. Cependant l'analyse que nous proposons pourra servir de support lors d'un éventuel développement sur station.

CONCLUSION

C'est dans les délais qui nous étaient impartis que nous avons menés cette étude à terme : nous avons réalisé l'étude préalable et les spécifications détaillées du futur logiciel d'ajustements statistiques de l'ORSTOM.

Ce projet nous a permis d'approfondir nos connaissances en statistique inférentielle et de nous familiariser avec les techniques d'ajustements. Cependant, nous retiendrons plus particulièrement l'intérêt de notre collaboration avec des experts en hydrologie dans la réalisation d'un logiciel de statistique.

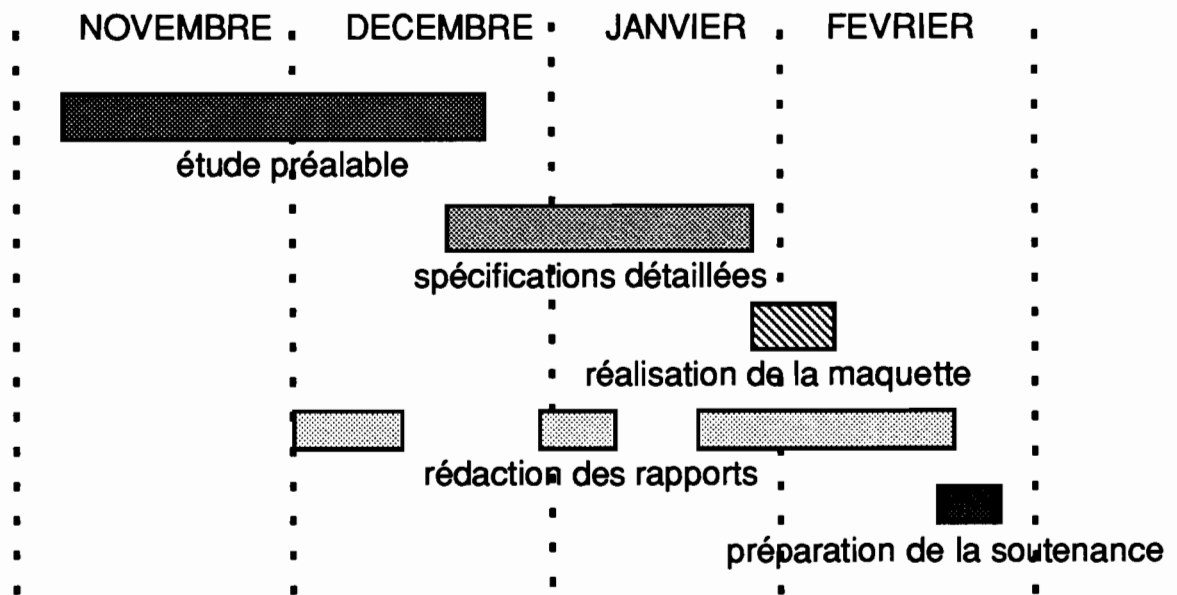
La confiance qu'ils nous ont accordés quant aux moyens et aux choix pour atteindre une solution répondant à leurs besoins, a été, pour nous une motivation supplémentaire

Nous avons conçu le futur logiciel de façon à ce qu'il soit ergonomique, simple d'utilisation et au service de l'hydrologue.

La réalisation est imminente. Nous souhaitons qu'il sera un outil de recherche efficace et qu'il pourra être diffusé à l'ensemble des hydrologues travaillant sur les ajustements statistiques.

De plus, nous souhaitons que notre collaboration avec les ingénieurs de l'ORSTOM ait été aussi satisfaisante pour eux qu'elle a été enrichissante pour nous. Nous espérons que cela concourra à entretenir de bonnes relations entre l'ORSTOM et les étudiants du département informatique et gestion de l'ISIM.

ECHEANCIER



BIBLIOGRAPHIE

Etude d'évaluation de logiciels d'ajustement de lois statistiques à des variables hydrologiques

H.LUBES J.M.MASSON Sept 1991.
Laboratoire d'hydrologie et de modélisation de Montpellier II.
Laboratoire d'hydrologie de l'ORSTOM.

Logiciel TROPHEE

(TRaitement des Observations PLuviométriques et Hydrologiques des Evènements Extrêmes).

Société BCEOM.

DIXLOI

(un ensemble de programme FORTRAN77 pour l'ajustement de lois statistiques et leurs représentations graphiques).

T. LEBEL J.F. BOYER ORSTOM Juin 1989.

Logiciel ALED

(Ajustement de Lois statistiques à des Echantillons de Données).

E. DELAFOSSE Oct 1989-Jan 1990
Laboratoire d'hydrologie et de modélisation de Montpellier II.

Logiciel d'hydrologie statistique

M. RISSON ISIM STE3.
Partie I Programmation.
Partie II Exemples d'exécutions.
Partie III Documentations.