OPIC
Office de la propriété intellectuelle du Canada

CIPO
Canadian Intellectual Property Office

(12) (19) (CA) **Demande-Application**

(21) (A1) **2,323,878**
(86) 1998/09/30
(87) 1999/09/23

(72) BOSSART, ROMAN, CH
(72) GRABINSKI, JORG, CH
(71) BÜHLER AG, CH

(51) Int.Cl.[6] G01N 21/27
(30) 1998/03/13 (198 10 917.2) DE

(54) **PROCEDE D'ETALONNAGE AUTOMATIQUE**
(54) **AUTOMATIC CALIBRATION METHOD**

(57) L'invention concerne un procédé d'étalonnage automatique relatif à des spectres d'un spectromètre en vue de l'analyse de substances dans l'industrie pharmaceutique, chimique, cosmétique, des colorants, des plastiques, du caoutchouc et des produits alimentaires.

(57) The invention relates to an automatic calibration method related to spectra of a spectrometer for examining substances in the pharmaceutical, chemical, cosmetic, dye, plastics, rubber, and foodstuff industries.

**PCT**

WELTORGANISATION FÜR GEISTIGES EIGENTUM
Internationales Büro

INTERNATIONALE ANMELDUNG VERÖFFENTLICHT NACH DEM VERTRAG ÜBER DIE INTERNATIONALE ZUSAMMENARBEIT AUF DEM GEBIET DES PATENTWESENS (PCT)

(51) Internationale Patentklassifikation 6 : G01N 21/27

A1

(11) Internationale Veröffentlichungsnummer: **WO 99/47909**

(43) Internationales Veröffentlichungsdatum: 23. September 1999 (23.09.99)

(21) Internationales Aktenzeichen: PCT/CH98/00418

(22) Internationales Anmeldedatum: 30. September 1998 (30.09.98)

(30) Prioritätsdaten:
198 10 917.2 13. März 1998 (13.03.98) DE

(71) Anmelder *(für alle Bestimmungsstaaten ausser US)*: BÜHLER AG [CH/CH]; CH–9240 Uzwil (CH).

(72) Erfinder; und
(75) Erfinder/Anmelder *(nur für US)*: BOSSART, Roman [CH/CH]; Wilerstrasse 89, CH–9230 Flawil (CH). GRABINSKI, Jörg [CH/CH]; Moosmatten 3b, CH–9244 Niederuzwil (CH).

(74) Gemeinsamer Vertreter: BÜHLER AG; Patentabteilung, CH–9240 Uzwil (CH).

(81) Bestimmungsstaaten: AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, TJ, TM, TR, TT, UA, UG, US, UZ, VN, ARIPO Patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), eurasisches Patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), europäisches Patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI Patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

**Veröffentlicht**
*Mit internationalem Recherchenbericht.*

(54) Title: AUTOMATIC CALIBRATION METHOD

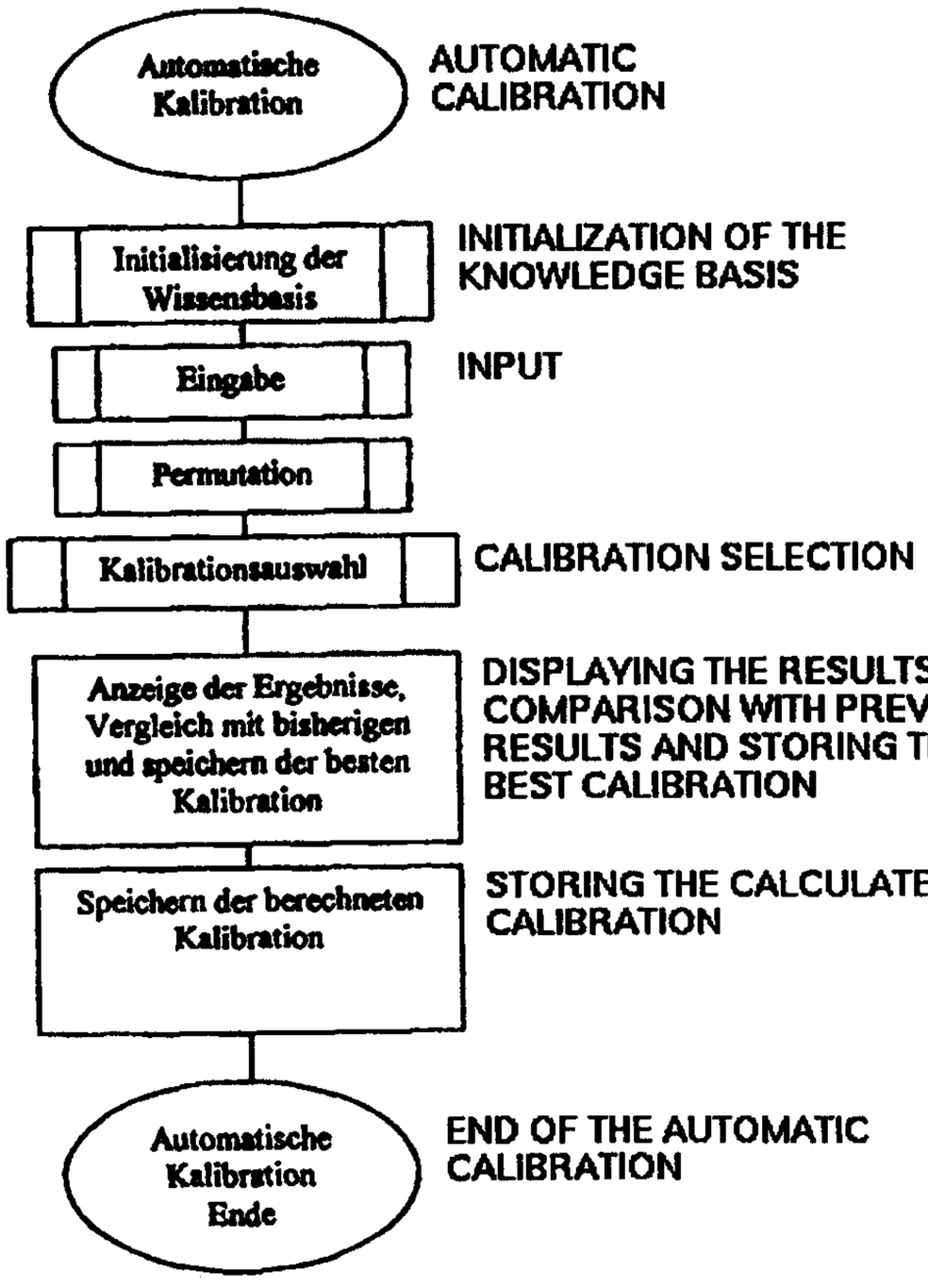
(54) Bezeichnung: AUTOMATISCHES KALIBRATIONSVERFAHREN

(57) Abstract

The invention relates to an automatic calibration method related to spectra of a spectrometer for examining substances in the pharmaceutical, chemical, cosmetic, dye, plastics, rubber, and foodstuff industries.

(57) Zusammenfassung

Die Erfindung betrifft ein automatisches Kalibrationsverfahren in Verbindung mit Spektren eines Spektrometers zur Untersuchung von Stoffen in der pharmazeutischen, chemischen, kosmetischen, Farb–, Kunststoff–, Gummi– und Nahrungsmittelindustrie.



**An automatic calibration method**

The invention relates to a calibration method for evaluating spectra of a spectrometer for examining solid, liquid or gaseous substances.

This calibration method serves the analysis of spectra produced by way of a spectrometer. Such analysis methods and devices function according to the principle whereby a transmitter emits electromagnetic radiation onto the substance to be examined and these rays which are reflected by or penetrate through the substance are recorded by a receiver. For these examinations all wavelength region of the electromagnetic spectrum may be used. Particularly suitable is the wave region in the near infra-red (NIR). Such methods and devices are amongst others mentioned in CH 685807 A5 or CH 683713 A5.

This analysis method is applied in particular in the pharmaceutical, chemical, cosmetic, colouring, plastics, rubber and foodstuffs industries.

Since with this spectral method there are various instrument types, sensors, substance tyes, aggregate conditions, mathematical methods and calibration behaviour and for each substance to be examined there are nominal spectra, before the actual evaluation a calibration is required.

Until now the calibration was effected by time-consuming, empirical, intuitive or "guessing" determination of the required variables cited above.

The object of the present invention then lies in specifying a calibration method which automatically in as short a time as

possible ascertains the optimal boundary conditions of the actual evaluation method.

According to the invention this object is achieved by the following method steps:

- defining and determining the number of components of a vector-equivalent data set
- producing the data sets on account of experience values, measuring methods, computation methods and/or specific substance properties
- acquiring nominal spectra
- calibration computation
- selecting primary and/or secondary factors
- determining and subsequent sorting of the calibration quality, in particular Q-value
- selecting the calibration on account of the best Q-values

wherein the calibration is effected automatically.

Various selection criteria (primary and secondary factors) are computed independently from the used calibration method (MLR/PCR/PLS/Cluster), wherein the data necessary for evaluation is processed.

Finally a quality factor optimal for the calibration is determined which results with the computation of the calibration method (MLR/PCR/PLS/Cluster).

Furthermore with the calibration method according to the invention the structure of the knowledge base which contains all data expert knowledge and their weightings may be developed.

Further advantageous method steps are deduced from the dependent claims. In particular as components of a data set the measuring device, in particular the sensor type, specifically optical and/or mechanical as well as the number of measurement run-throughs may be ascertained. As further components of the data sets necessary for the calibration the specific data of the substance to be examined is determined. Furthermore likewise in the form of components of a data set the computation method and the behaviour of the calibration may be layed down. As computation methods for the calibration in particular the methods: MLR, PCR, PLS and/or CLU known in chemometry are used.

For calibration further there serve the nominal spectra known for a substance to be examined.

In order to obtain an optimal calibration, the components of the data sets completely and/or partly may be permutated within the calibration method.

The measured spectra are generally characterised by a multitude of oscillation superpositions. A visual evaluation is therefore practically not possible. Specific differences in the spectra of similiar substances often consist only of a slight displacement or small shape changes of the mostly wide absorption bands. It is therefore necessary to evaluate these spectra with mathematical methods. These mathematical methods are based on chemometric software. Chemometry is to be understood as the application of mathematical methods in chemistry. The chemometric software in spectroscopy has the exercise of finding a statistical relationship between spectral data and the known property values of the samples used for calibration.

As an example, in the following, the invention is represented in the form of flow diagrams. There are shown in:

Figure 1 schematic representation of a possible measuring device

Figure 2 a flow diagram of the automatic calibration

Figure 3 initialisation of the knowledge base

Figure 4 input

Figure 5 permutation

Figure 6 method computation

Figure 7 factor selection

Figure 8 qualitative and quantitative Q-value determination

In Figure 1 there is represented a possible measuring arrangement. The samples 6 are acquired by a measuring head 5 containing a sensor 5, wherein a spectrometer 4 records the spectra. The digitalised data are via a lead 3 led into a computer 1 which then carries out the calibration according to the invention. The spectra and the evaluation is usually displayed on a monitor 2.

In Figure 2 the basic course of the automatic calibration according to the invention is shown. The first step comprises the initialisation of the knowledge base. There follows the actual input of the data with the formation of data sets. The components of the data sets are completely or partly permutated and

are subjected to a computing method. The used permutations are initialised, wherein for each permutation a special factor selection, a Q-value is computed and thus a specific calibration is allocated and stored. There results therefrom a multitude of quality values from which the optimal Q-value may be selected and thus the most suitable calibration may be determined.

Figure 3 represents the initialisation of the knowledge base. The knowledge base forms the basis for the components containing the data sets. To the initial data there belong for example the instrument type/sensor, the substance type, the calibration type, the calibration behaviour and/or spectral data. From these five initial data, on account of expert knowledge the parameter group is determined and weighted. As a sensor type an optical or mechanical sensor may be considered. The substance type is layed down by way of a prior treatment of data and by way of the selection of the measuring procedures. To the calibration type there belongs a calibration data set with a possible computing method. The calibration behaviour is determined from the selected wavelength region, a pre-treatment of data, the primary, the secondary factors and/or the number of measuring procedures. The spectra data are composed of the calibration data set, the validation data set, the wavelength selection, the pretreatment of data and/or the number of measuring procedures.

After the initialisation there follows the actual input according to Figure 4. From the spectra data there is selected in particular the instrument type/sensor with which the respective spectrum is recorded. Furthermore the substance type, the calibration type and the calibration behaviour are inputted.

After the input there follows according to Figure 5 the complete and/or part permutation of the components of the data sets. The

respective parameter group is subjected to a computation for the calibration. With this, the method itself forms a component of the data set. The method computation is broken down in Figure 6.

The computation course begins with the main component analysis, amongst others published in S. Wood, K. Esbensen, P. Geladi, "Principle Component Analysis", Chemometrics and Intelligent Laboratory Systems 2 (1987) 37- 52. A spectrum measured with the device according to Figure 1 may for example be composed of for example 500 measuring points. This individual measuring data corresponds to the intensity values in dependence on 500 wave numbers in the near infra-red. In order to obtain a good calibration, one in turn requires a large number of spectra. For example 100 substance spectra thus result in 50,000 data points which entails a great computing effort. In order to obtain acceptable computing times the spectral data is condensed with the help of the main component analysis. With this no important information is lost.

With a calibration, to the samples there are allocated two different data sets independent of one another, specifically the calibration data set on the one hand and the validation data set on the other hand. With the calibration data set the main component analysis is carried out. With the validation data set the results of the calibration are evaluated.

For the quantitative calibration there are available a multitude of computation methods. In particular there are cited here the three most common computation methods: Multiple Linear Regression (MLR), Principle Component Regression (PCR) and the Partial Least Squares Regression (PLS), for example published in Bruce R Kowalski, "Chemometrics, Mathematics & Statistics in Chemistry"

NATO AIS Series, Series C: Math. & Phys. Sciences Vol. 138 (1984).

MLR is an extension of the linear regression to several dimensions. This evaluation on account of a few selective wavelengths requires no main component analysis. In this method the properties are computed by intensity values and correlation coefficients.

The PCR is composed of two steps. In the first step the intensities to be loaded are computed by way of the main component analysis. The second step produces the correlation coefficients by way of MLR.

According to the principle of recursion with PLS the data to be loaded is computed. The quantitative reference values are however taken account of already at the beginning of the computation. Whilst the PCR reduces the spectral data to the most dominant dimensions, the PLS targets the most relative dimensions, i.e. the best agreement between prognosed and actual values.

The actual calibration is made up on the one hand of the qualitative and on the other hand of the quantitative calibration.

The course of the quantitative calibration begins with the measuring of the calibration spectra. In order to set up a representative calibration a multitude of various charges of the same class should be measured. With an increasing number of measuring run-throughs the signal/optical noise ratio may be improved and inhomogenities compensated.

After the measuring of the samples, the actual setting up of the calibration begins.

1. The spectra are divided into a calibration data set and into a validation data set, wherein the selection is effected purely randomly. As mentioned above, the calibration data set must be independent from the validation data set. In this first set one works without a wavelength selection and without data pretreatment.

2. The main component analysis is carried out.

3. The total number of the primary factors is selected. Primary factors are to be understood as the factors which are meaningful for the description of the spectra up to the optical noise limit of the applied measuring method.

4. Following the detemination of the primary factors, the selection factors, called secondary factors are selected. These selection factors are the factors which allocate to the associated spectra an unequivocal separation of the various calibration qualities. The factor selection is shown in the form of a flow diagram in Figure 7.

5. There follows an optimisation of the calibration. If no selection is achieved the various data pretreatments should be carried out. Those wavelength regions are rejected which do not contain any significant information.

6. The best calibration is then stored and tested.

The qualitative calibration is used for measuring property values (e.g. water content, mixing consituent parts, hydroxy number, etc.).

After measuring the samples the calibration is set up. It is effected essentially according to the same course as with the qualitative calibration.

According to the type of calibration, qualitative or quantitative, according to the flow diagram in Figure 8, the quality of the calibration, the so-called Q-value, is layed down.

At the end of the calibration computations subsequently a table of the various Q-values is set up from which one may then select the optimal calibration.

**Patent claims**

1. A calibration method for evaluating spectra of a spectrometer for examining solid, liquid or gaseous substances characterised by the following method steps:

   - defining and determining the number of components of a vector-equivalent data set
   - producing the data sets on account of experience values, measuring methods, computation methods and/or specific substance properties
   - acquiring nominal spectra
   - calibration computation
   - selecting primary and/or secondary factors
   - determining and subsequent sorting of the calibration quality, in particular Q-value
   - selecting the calibration on account of the best Q-values

   wherein the calibration is effected automatically.

2. A method according to claim 1, characterised in that as components of a data set the measuring device, in particular the sensor type, specifically optical and/or mechanical, the wavelength region, the adaptation of the required data as well as the number of measuring run-throughs are layed down.

3. A method according to claim 1 or 2, characterised in that as components the specific data of the substance to be examined is determined in a processed manner.

4. A method according to at least one of the claims 1, 2 or 3, characterised in that as a component the computation method is layed down.

5. A method according to claim 4, characterised in that as computation methods for the calibration the methods: MLR, PCR, PLS and/or CLU known per se in chemometry are provided.

6. A method according to at least one of the preceding claims, characterised in that as a component the behaviour of the calibration is provided.

7. A method according to at least one of the preceding claims, characterised in that the spectra of the substance to be examined serve as components.

8. A method according to at least one of the preceding claims, characterised in that the components of the data sets for optimising the calibration are completely and/or partly permutated.

9. A method according to at least one of the preceding claims, characterised in that the optimal number of primary factors for describing the spectra is determined up to the optical noise limit of the applied measuring method.

10. A method according to at least one of the preceding claims, characterised in that after the determining of the number of primary factors the selection factors described as secondary factors are determined which permit an unequivocal separation of the various Q-values of the calibrations.

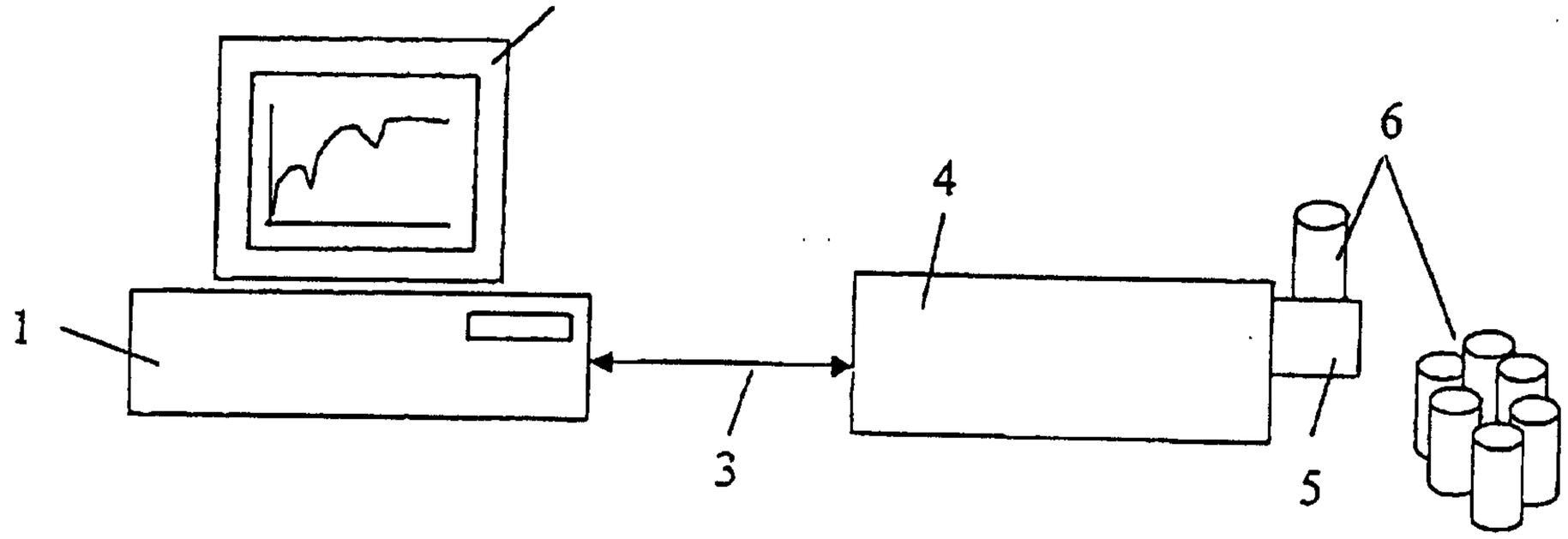
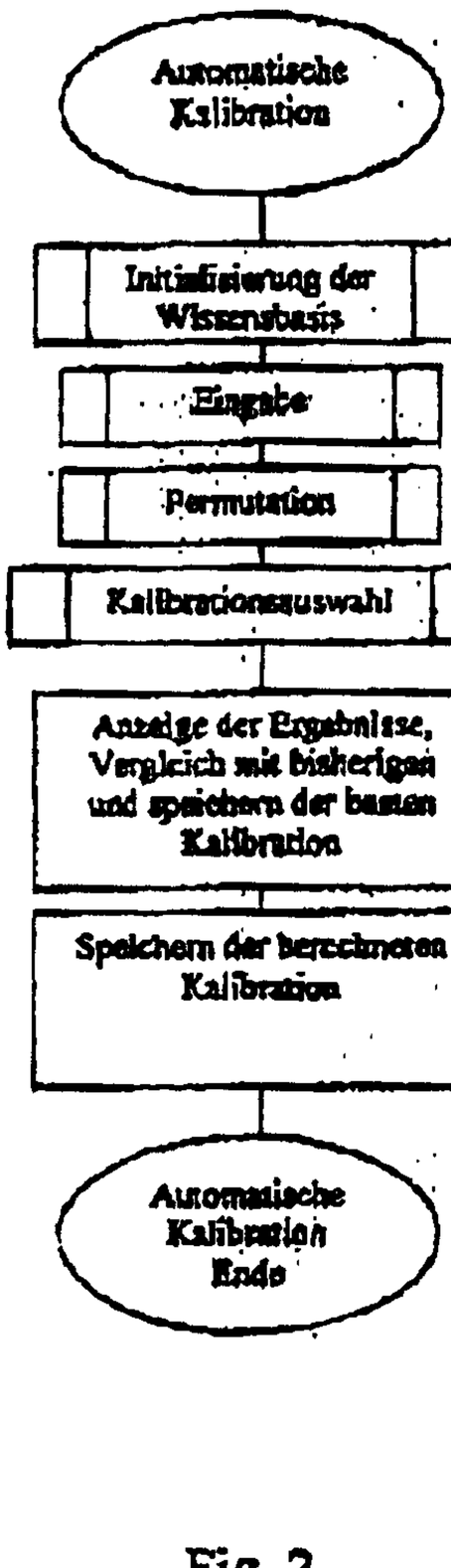



Fig. 1




automatic calibration

initialisation of the knowledge base

input

permutation

calibration selection

display of results comparison with previous ones and storage of the best calibration

Storage of the computed calibration

end of automatic calibration

Fig. 2

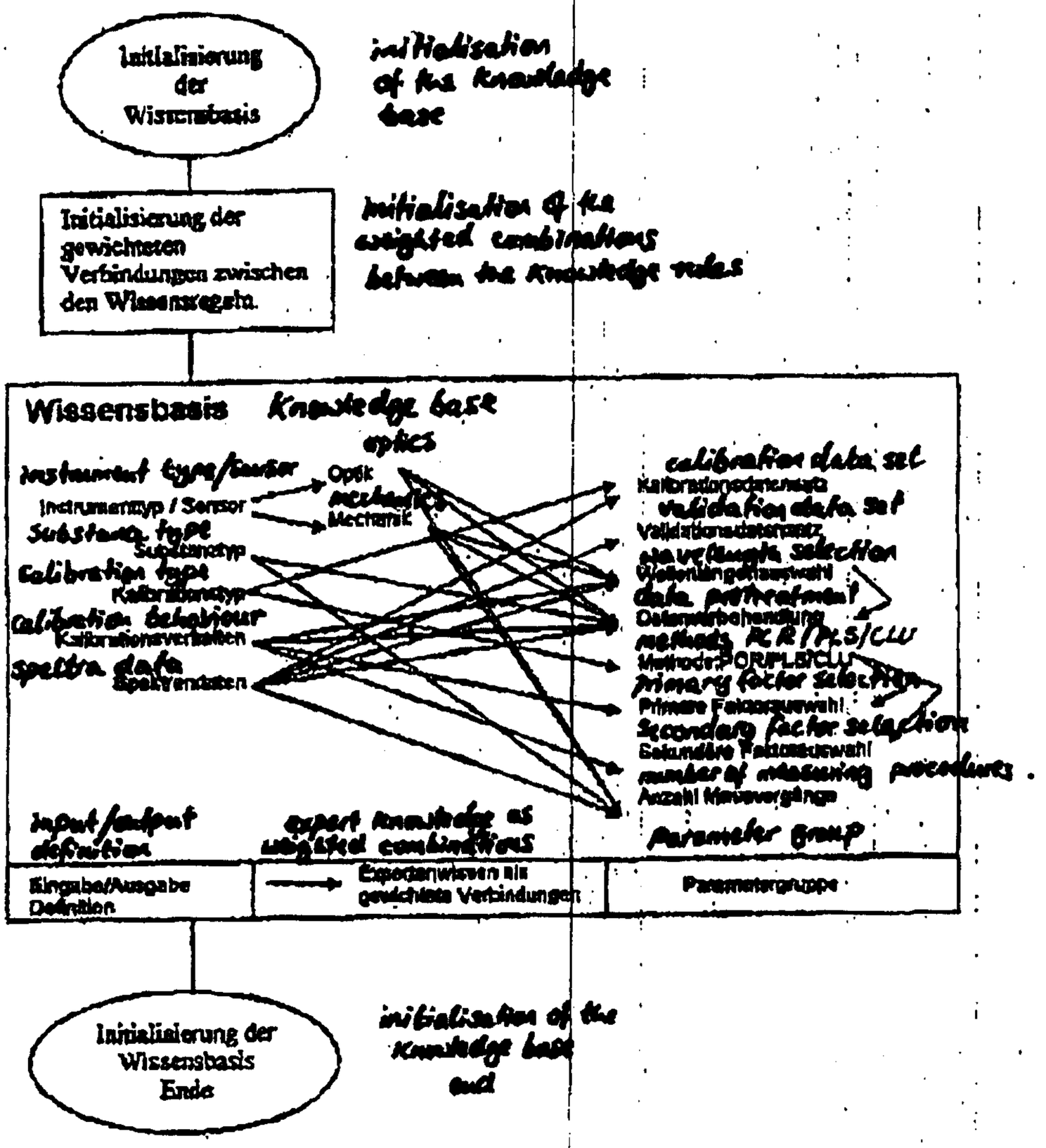

Initialisierung der Wissensbasis
initialisation of the knowledge base
Initialisierung der gewichteten Verbindungen zwischen den Wissensregeln
initialisation of the weighted combinations between the knowledge rules
Wissensbasis
Knowledge base
optics
instrument type/sensor
Optik
Instrumententyp / Sensor
Mechanik
substance type
Substanztyp
calibration type
Kalibriertyp
calibration behaviour
Kalibrationsverhalten
spectra data
Spektrendaten
calibration data set
Kalibrationsdatensatz
validation data set
Validationsdatensatz
data preselection
Datenvorbehandlung
method: PCR/PLS/CLV
Methode: PCR/PLS/CLV
primary factor selection
Primäre Faktorauswahl
secondary factor selection
Sekundäre Faktorauswahl
number of measuring procedures
Anzahl Messvorgänge
input/output definition
expert knowledge as weighted combinations
parameter group
Eingabe/Ausgabe Definition
Expertenwissen als gewichtete Verbindungen
Parametergruppe
Initialisierung der Wissensbasis Ende
initialisation of the knowledge base end

Fig. 3

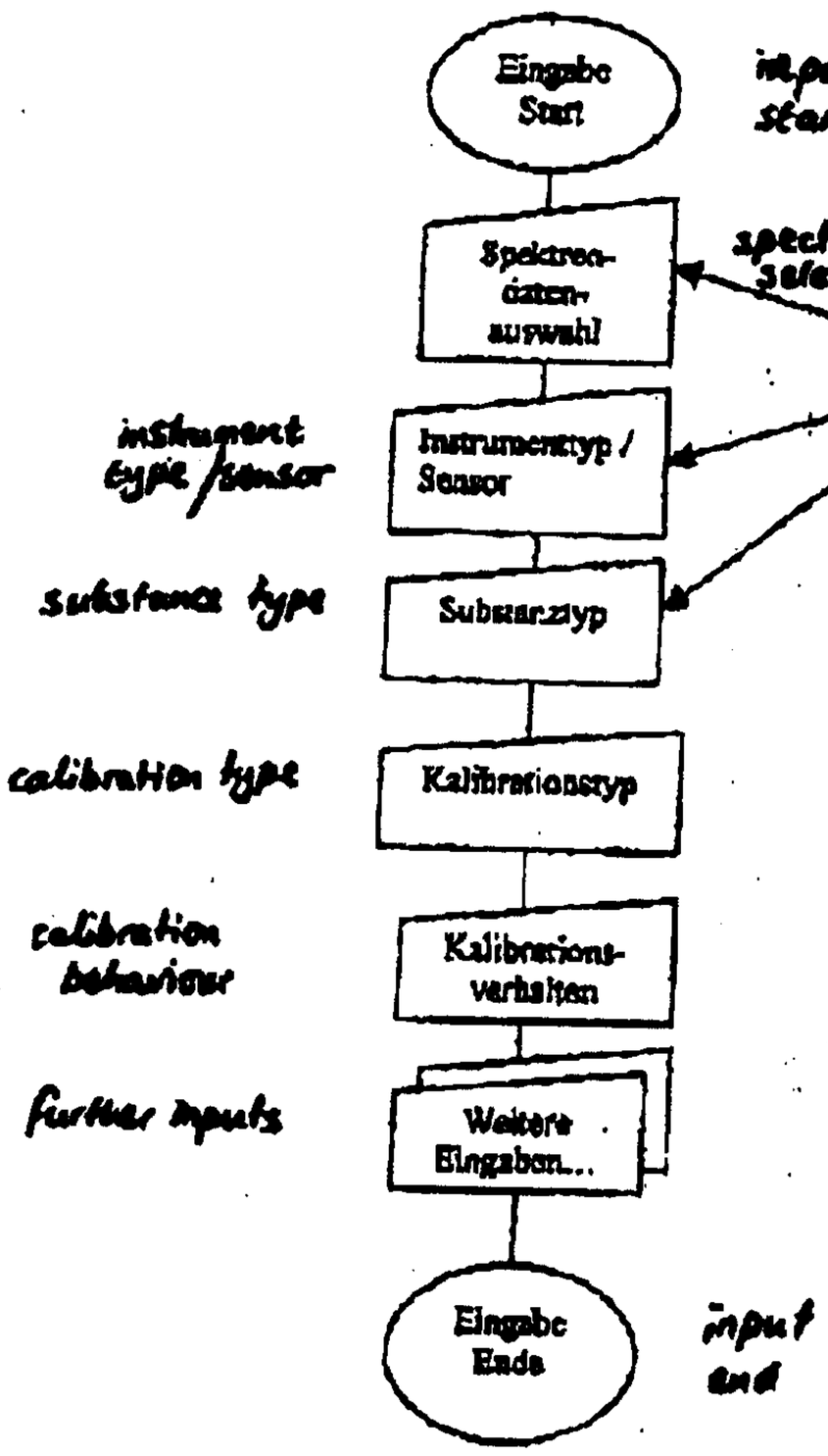



Fig. 4

McFadden, Fincham

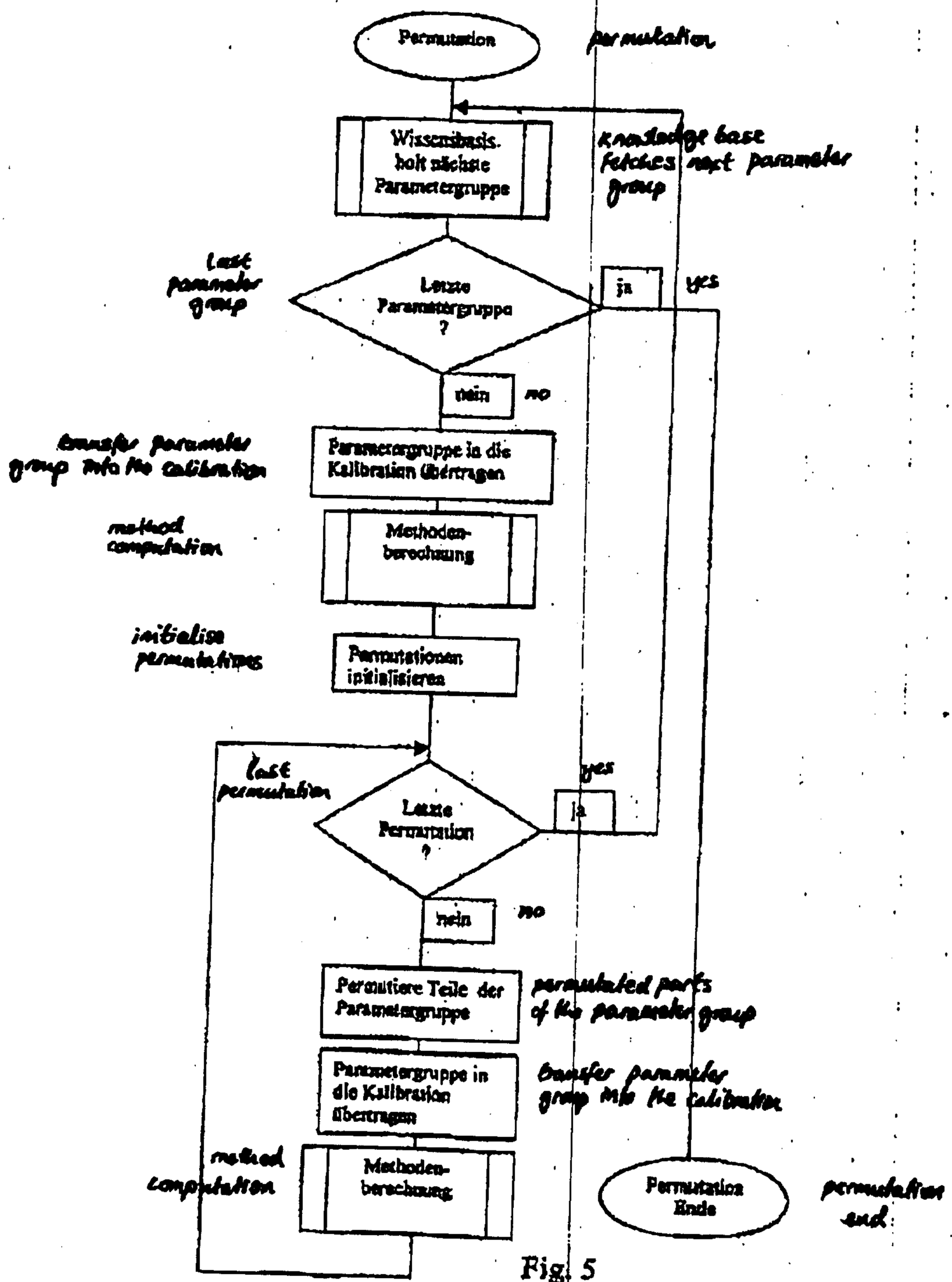



Fig. 5

McFadden, Fincham

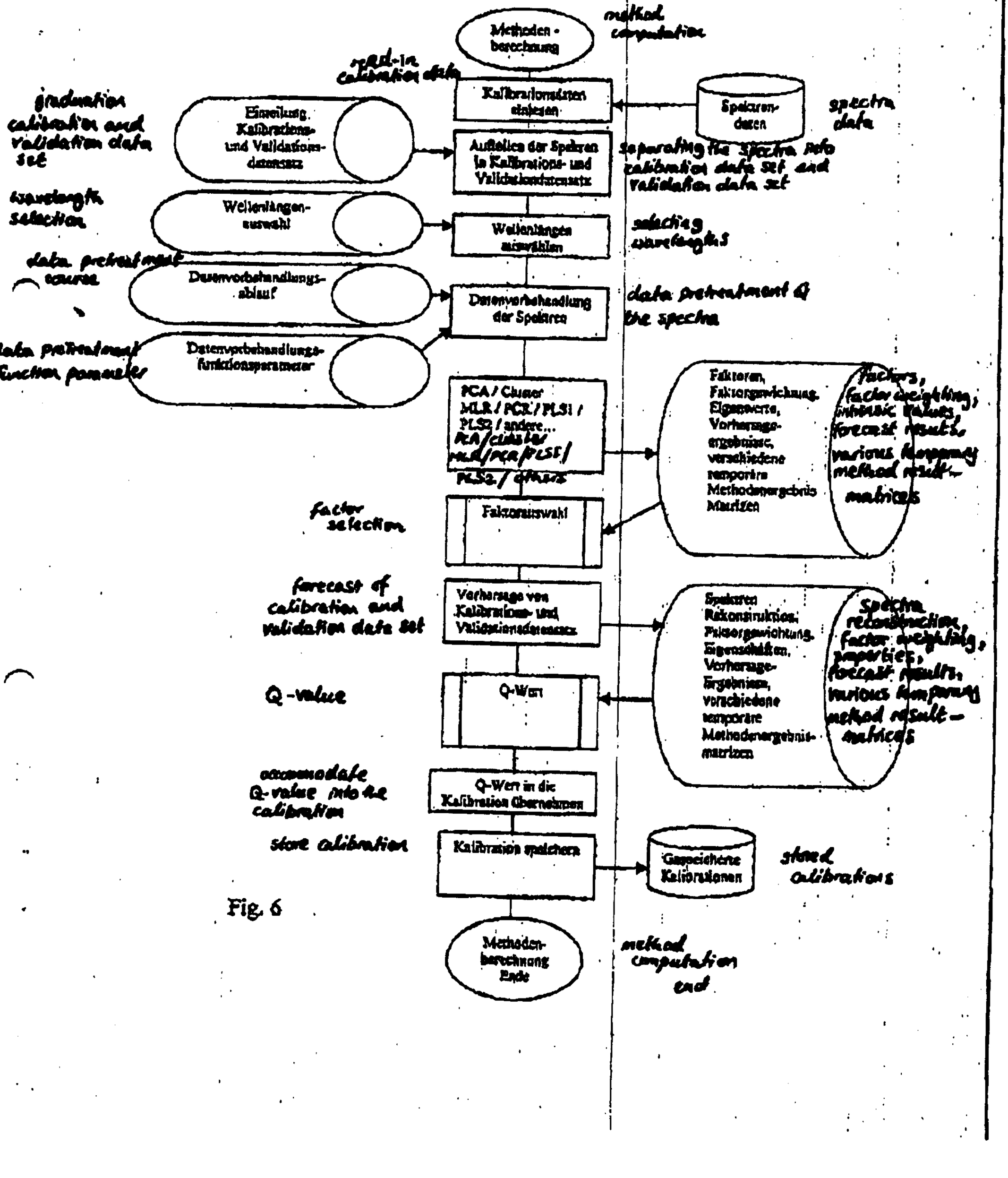

Methoden-berechnung
Kalibrationsdaten einlesen
Spektrendaten
Einteilung Kalibrations- und Validations-datensatz
Aufteilen der Spektren in Kalibrations- und Validationsdatensatz
Wellenlängen-auswahl
Wellenlängen auswählen
Datenvorbehandlungs-ablauf
Datenvorbehandlung der Spektren
Datenvorbehandlungs-funktionsparameter
PCA / Cluster MLR / PCR / PLS1 / PLS2 / andere...
Faktoren, Faktorgewichtung, Eigenwerte, Vorhersage-ergebnisse, verschiedene temporäre Methodenergebnis Matrizen
Faktorauswahl
Vorhersage von Kalibrations- und Validationsdatensatz
Spektren Rekonstruktion, Faktorgewichtung, Eigenwerten, Vorhersage-Ergebnissen, verschiedene temporäre Methodenergebnis-matrizen
Q-Wert
Q-Wert in die Kalibration übernehmen
Kalibration speichern
Gespeicherte Kalibrationen
Methoden-berechnung Ende

Fig. 6

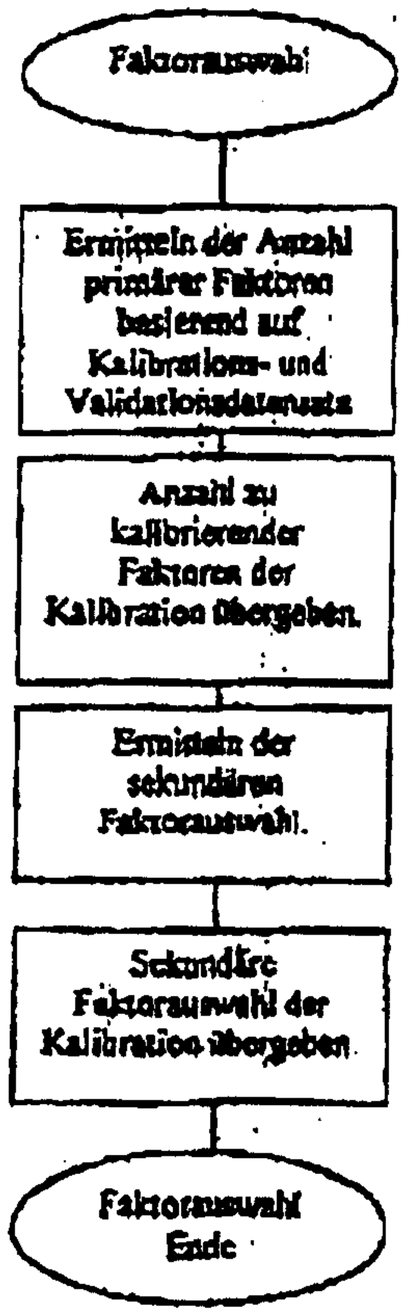



Fig. 7

factor selection

determining the number of primary factors based on the calibration and validation data set

transfer the number of factors to be calibrated to the calibration

determining the secondary factor selection

transfer the secondary factor selection to the calibration

factor selection end

McFadden, Fincham

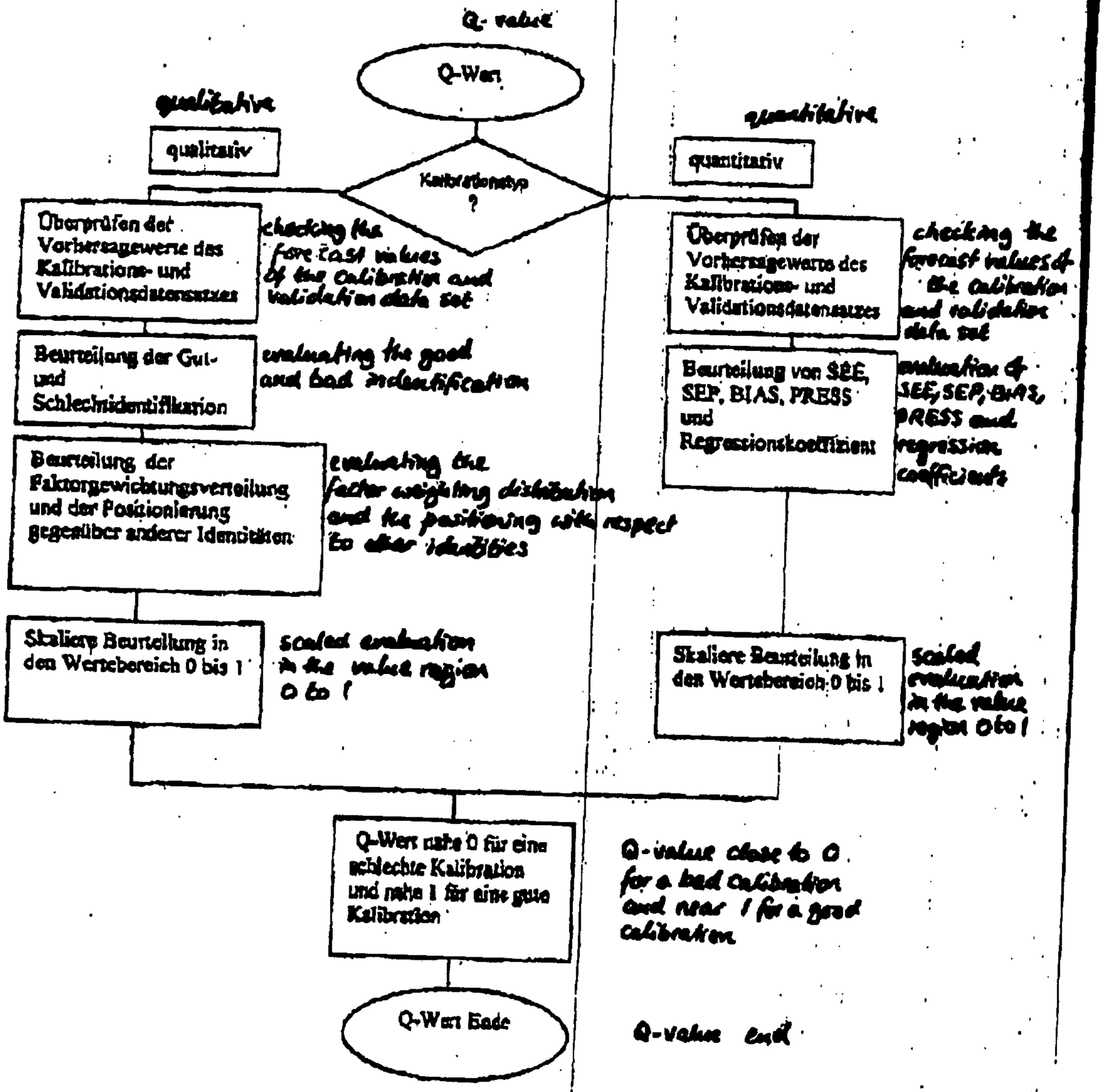



Fig. 8

McFadden, Fincham