Chemometric analysis of acid-base measurements

A multivariate approach

Marcel Hekking

Acknowledgement

The work presented in this thesis was partly supported by a grant for scientific research from the 'Onze Lieve Vrouwe Gasthuis' (Amsterdam, The Netherlands) and the 'St. Elisabeth Ziekenhuis' (Tilburg, The Netherlands).

Financial support by IMRO TRAMARKO bv for the publication of this thesis is gratefully acknowledged.

ISBN: 90-9013351-8

Chemometric analysis of acid-base measurements

A multivariate approach

Chemometrische analyse van zuur-base metingen

Een multivariate benadering

PROEFSCHRIFT

ter verkrijging van de graad van doctor aan de Erasmus Universiteit Rotterdam op gezag van de Rector Magnificus Prof.dr. P.W.C. Akkermans M.A. en volgens besluit van het College voor Promoties.

De openbare verdediging zal plaatsvinden op

woensdag 22 december 1999 om 15:45 uur

door

Marcel Hekking

geboren te 's-Gravenhage

Promotiecommissie

Promotor	Prof.dr. E.S. Gelsema
Overige leden:	Prof.dr. H.A. Bruining, Prof.dr.ir. H.L. Vader en Prof.dr.ir. A. Hasman
Co-promotor	Dr. J. Lindemans

THE FAR SIDE by Gary Larson

Table of Contents

Chapter 1

General introduction

1.1 The role of clinical chemistry in medicine

Medicine is an art and a science in the service of fellow human beings [1]. On the basis of collected empirical data and information, clinicians select specific diagnoses, rule out other differential diagnoses and eventually make decisions about which and how specific therapeutic interventions are made for the benefit and health of their patients. For a proper interpretation, collected data and information must be compared with other, already existing, data and information to assess the exact value of the clinician's findings. Moreover, a clinician compares observed medical data of a patient with knowledge obtained during his or her training as a clinician and with the experience obtained by working with other patients.

A prerequisite in this paradigm, however, is that collected empirical data on which the diagnoses of a clinician are based must be as objective as possible. Clinical chemistry takes a pivotal role in this in the sense that the chemical characterisation of a patient's body fluid is one of the ways in medicine that can provide such objective data. Since the beginning of this century, clinical chemistry has evolved into a separate and independent discipline in the field of medicine [2-4]. Nowadays, most often a single central clinical chemistry laboratory takes care of the 'analytical needs' of one or more hospitals.

Tasks of the clinical chemist typically include the improvement of existing methods of chemical analysis, the development of new analytical methods and providing the clinician with as much information as possible on the basis of chemical analyses. Especially this last task forms the basis of what has

become known as *chemometrics*, a branch of clinical chemistry that uses mathematical and statistical methods to extract a maximum of information from chemical analyses [5, 6].

This thesis presents a multivariate chemometric approach to the problems that are currently associated with the interpretation and evaluation of those laboratory measurements that are used to assess the arterial acid-base status of a patient in an intensive care unit (ICU).

1.2 Arterial acid-base measurements in the ICU

The ICU of today is a highly specialised ward in which expert medical, nursing and technical staff provides medical services to severely ill patients. It is characterised as a high-tech environment in which the real-time monitoring of vital functions plays a central role. The origin of the ICU can be traced back to the second half of the 19^{th} century when special rooms, adjacent to the operating room, were used primarily for the purpose of postoperative care [7]. In the course of time, these recovery rooms evolved into specialised respiratory care units and shock and trauma units, eventually leading to the present day ICU. The modern ICU provides integrated cardiopulmonary support for both medical and surgical patients suffering from severe respiratory and / or cardiac problems as a result of disease or trauma.

The most frequently ordered chemical test in the ICU is the arterial blood gas measurement [8]. Arterial blood gas measurements comprise those measurements of the patient's arterial blood that are used for the evaluation and interpretation of the patient's oxygen and acid-base status. Basic arterial blood gas measurements include: the partial pressure of oxygen (PaO_2), the oxygen saturation of haemoglobin, the pH of arterial blood, the partial pressure of carbon dioxide ($PaCO_2$) and the bicarbonate-ion concentration ($a[HCO_3^-]$). The first two measurements (PaO_2 and oxygen saturation) are used to evaluate the oxygen status, while the other three are used for the interpretation of the arterial acid-base status.

In a strict sense, the term blood gas measurements is incorrect, since only PaO_2 and $PaCO_2$ are true gas measurements and in modern chemical analysers $a[HCO_3^-]$ is not measured but calculated from measured pH and $PaCO_2$. Moreover, two other derived acid-base parameters are generally

considered part of the set of arterial blood gas measurements. These parameters are the standard bicarbonate-ion concentration (SB) and the base excess (BE). Their derivation and rationale are described in section 1.4.3 in more detail.

Since the second half of this century, the analysis of arterial blood for the purpose of acid-base characterisation has become a vital part of intensive care medicine. The importance of the acid-base characterisation of arterial blood is illustrated by the severe polio epidemic that struck Copenhagen (Denmark) in 1952 [9]. During this epidemic, hospitals in Copenhagen had to cope with a large number of patients needing intensive artificial respiration as a result of paralysis of the respiratory muscles. For a proper setting of the artificial respiration, the complete acid-base status of the patient had to be known. At that time, arterial blood of patients was seldom sampled for the purpose of performing blood gas measurements [10]. Arterial blood gas measurements were mainly performed in physiological laboratories and were not part of daily clinical practice. Techniques of measurement were cumbersome and needed large equipment.

The clinical necessity of quickly knowing the patient's arterial acid-base status for the purpose of a proper adjustment of the artificial respiration inspired Poul Astrup to develop his equilibration method [9]. This method allowed a relatively quick determination of the three basic acid-base parameters by only measuring the pH of an arterial blood sample and the pH of the sample equilibrated at two known $PaCO_2$ gas tensions. The original $PaCO_2$ is calculated by interpolation [11, 12]. Since then, techniques of analysis developed and arterial acid-base measurements have become routine and indispensable in the daily clinical care of intensive care patients.

1.3 Basic acid-base physiology

In chemical terms, acids are substances that are capable of donating hydrogen (H^+) ions while bases are substances capable of accepting H^+ ions. The amount of H^+ ions in the arterial blood determines its actual acidity. Acidity is measured as pH, which is, according to the definition of Sörensen, the negative logarithm of the H^+ concentration ($[H^+]$) [9].

The regulation of the amount of H^+ ions in the arterial blood and conse-

quently its pH is one of the most powerful controlling mechanisms in the human body. Under normal physiologic conditions, the pH of arterial blood is kept within well-defined limits. This tight regulation of the H^+ concentration in arterial blood is essential since H^+ ions are highly reactive with negatively charged parts of molecules. Changes in H^+ concentration (intracellular as well as extra-cellular) therefore have a profound influence on the molecular configuration and consequently on protein function [13]. Hence, maintaining a constant pH ensures an optimal working condition for enzymes and other proteins. Moreover, large deviations in pH may have effects on the nervous system. If the body becomes too acidic, the nervous system can become so depressed that death can occur. On the other hand, if the body becomes too alkaline, the nervous system can become overexcited, resulting in death from tetanus of the respiratory muscle [14].

Two mechanisms exist to regulate pH of arterial blood: long term physiological buffering and short term chemical buffering. Physiological buffering is the redistribution, production, excretion and/or retention of (non-)volatile acids and bases by means of physiological processes. Chemical buffering is the result of the presence of weak acids and their conjugated bases in the arterial blood. Examples of chemical buffers in arterial blood are: inorganic phosphate, organic phosphate and haemoglobin.

One of the most important chemical buffer systems in the blood, however, is the bicarbonate ion (HCO_3^-)/carbon dioxide (CO_2) buffer system. It is mainly the presence of this buffer system that makes it possible for the human body to cope with the constant load of exogenous acids and bases and the vast amount of both volatile and non-volatile acids that are continuously generated as a result of normal metabolism.

The equation describing the HCO_3^-/CO_2 buffer system in blood is:

$$CO_2 + H_2O \rightleftharpoons H_2CO_3 \rightleftharpoons H^+ + HCO_3^- \tag{1–1}$$

The left-hand side of this chemical reaction represents the formation of carbonic acid (H_2CO_3) from CO_2 and H_2O. Therefore, although CO_2 itself is not an acid, an elevation of the CO_2 in the blood increases the acidity of the blood through the formation of H_2CO_3 which immediately dissociates into protons (H^+) and bicarbonate ions (HCO_3^-).

Since the concentration of H_2CO_3 is so low in relation to the concentration of dissolved CO_2 and the concentration of HCO_3^-, the law of mass action for the HCO_3^-/CO_2 buffer system is:

$$K = \frac{[H^+] \times [HCO_3^-]}{[CO_2] \times [H_2O]}$$

(1–2)

where K is a constant.

Because H_2O is relatively constant in body fluids, it can be omitted from the equation and incorporated into the constant K, further indicated as K′ [13]. Rewriting the resulting equation to solve $[H^+]$ yields the equation that Lawrence Joseph Henderson (1878-1942) first described in 1909 [10]:

$$[H^+] = K' \times \frac{[CO_2]}{[HCO_3^-]}$$

(1–3)

The concentration of dissolved CO_2 in blood ($[CO_2]$) is proportional to the partial pressure of CO_2 (PCO_2) in the gas with which the blood is in equilibrium. Therefore, $[CO_2]$ can be replaced by the partial pressure of CO_2 in the blood. Partial pressures are either measured in millimetres mercury (mmHg) or kilo-Pascal (kPa) where 1 mmHg = 0.133 kPa. The constant relating $[CO_2]$ in mmol/l to the PCO_2 is called the solubility constant. The solubility constant for $[CO_2]$ in plasma is 0.03 mmol per litre per mmHg or 0.225 mmol per litre per kPa.

Moreover, applying the pH concept of Sörensen, in 1917 Karl Albert Hasselbalch (1874-1962) introduced the Henderson-Hasselbalch equation:

$$pH = pK' + \log\frac{[HCO_3^-]}{PCO_2}$$

(1–4)

where $pK' = 6.10$ and α is the solubility constant for $[CO_2]$ in plasma.

From equation 1–4 it is apparent that pH is the resultant of the ratio $[HCO_3^-]/PCO_2$. Both PCO_2 and $[HCO_3^-]$ can effectively be regulated by lungs and kidneys, respectively [15]. This feature in particular makes

the HCO_3^-/CO_2 buffer system so effective in maintaining a constant arterial blood pH. Knowing pH, PCO_2 and $[HCO_3^-]$ in the arterial blood of a patient is vital when interpreting the acid-base status of arterial blood. It gives information on both the respiratory and metabolic component of an acid-base disturbance and their joint effect on the acidity of the arterial blood.

Although Sörensen introduced the electrochemical measurement of H^+ ions as early as in 1909, it was not until 1932 that pH glass electrodes were produced commercially and used on a regular basis. Before that time, pH of blood was indirectly obtained from measuring total CO_2 and PCO_2 in the blood with the manometric Van Slyke apparatus that Donald Dexter van Slyke (1883-1971) introduced in 1924 [9]. Around 1960 the CO_2 electrode was introduced into clinical chemistry. Today, chemical analysers measure pH and PCO_2 and calculate $[HCO_3^-]$ with the use of the Henderson-Hasselbalch equation (see Equation 1 –4).

1.4 The clinical interpretation of acid-base parameters

An impairment in either the respiratory or metabolic function (or both) of the body may result in so-called acid-base disturbances [13]. For a proper treatment of these disturbances it is essential for an ICU clinician to be aware of the exact acid-base status of the arterial blood of an ICU patient. With the analysis of measured and calculated arterial acid-base parameters, the ICU clinician aims to find the underlying cause(s) of one or more acid-base disturbances in order to remove it with specific therapeutic interventions. Moreover, for patients receiving artificial respiration, the acid-base analysis of arterial blood is essential for setting the kind and degree of artificial respiration.

1.4.1 General nomenclature and terminology

Acid-base disorders can be divided into *primary*, *secondary* and *combined* acid-base disturbances. Primary acid-base disturbances are the result of impairment of either the respiratory function or the metabolic function of the body. Impairments of the respiratory function result in primary *respiratory* acid-base disturbances, whereas impairments in metabolic function result in

non-respiratory or *metabolic* disturbances. Both respiratory and metabolic disturbances can be further divided into disturbances that tend to lower the pH, resulting in *acidemia*, and disturbances that tend to raise the pH, resulting in *alkalemia*. These acid-base disturbances are called *acidoses* and *alkaloses*, respectively. Hence, the terms acidosis and alkalosis refer to underlying pH-deranging physiologic processes, whereas the terms acidemia and alkalemia merely indicate the actual acidity of arterial blood. Multiple single primary acid-base disturbances can be present at the same time, resulting in *combined* acid-base disturbances.

Moreover, as a response to primary acid-base disorders, the human body is capable of initiating compensating mechanisms. Primary respiratory disturbances trigger mechanisms in the kidneys that actively regulate the reabsorbtion of excreted HCO_3^- ions, thereby inducing metabolic compensating effects. Also, primary metabolic dysfunction eventually triggers the breathing centre, resulting in an adjustment of the respiration and consequently the $PaCO_2$. These compensating processes result in *secondary* acid-base disturbances. The capability of the body to compensate for primary acid-base disturbances prevents large changes in the pH of arterial blood even though pathological processes may be present.

Respiratory compensations are very rapid and effective within minutes, while metabolic compensations can take up to three days to be fully effective. A metabolic compensation can, however, when in full working order, completely compensate a primary respiratory disturbance, while a respiratory compensation can only partially compensate primary metabolic acid-base disturbances.

It is apparent that for a proper treatment of an acid-base disturbance, the complete acid-base status of a patient should be known to a clinician. Although the body can compensate primary acid-base disturbances to a certain extent, therapeutic measurements must be taken as soon as possible to eliminate any primary acid-base disturbance. Moreover, severely ill patients on the ICU most often receive some form of artificial respiration. Being on mechanical ventilation means that the body cannot fully employ respiratory compensating mechanisms, making the ICU clinician even more responsible for keeping the pH of the arterial blood within acceptable boundaries.

For most ICU patients, an arterial blood gas analysis is performed on a routine basis, for instance every 3 or 6 hours. However, the interpretation of

acid-base data is still regarded as difficult since several pieces of information must be evaluated at the same time in their clinical context. Multiple primary disturbances can be present at the same time, concealed by various degrees of compensation, making the diagnosis and monitoring of acid-base data a complex task.

This complexity is illustrated by the coexistence of two distinct methods for interpreting arterial acid-base parameters. One method uses *in vivo* information to interpret pH, $PaCO_2$ and $[HCO_3^-]$, while the other method makes use of pH, $PaCO_2$ and a calculated *in vitro* parameter called base excess (BE). This latter method was developed around 1960 by Poul Astrup and Ole Siggaard-Andersen from Denmark and is therefore also known as the *Scandinavian view* [16].

Schwartz and Relman of the Tufts University School of Medicine in Boston (USA) criticised the *in vitro* approach and made a case for pH, $PaCO_2$ and $[HCO_3^-]$ [17]. This method is therefore also known as the *North American view*. The controversy between the two schools, which Bunker called 'The Great Trans-Atlantic Acid-Base Debate', still exists today, although many attempts were made to bridge the gap [16, 18-22].

1.4.2 The North American view; $[HCO_3^-]$ and in vivo CO_2 buffer lines

In the North American view, a high value of $PaCO_2$ indicates a primary respiratory acidosis or a respiratory compensation for a metabolic alkalosis, while a low value of $PaCO_2$ indicates a primary respiratory alkalosis or a respiratory compensation for a primary metabolic acidosis. The metabolic component of an acid-base status is assessed with $[HCO_3^-]$. A high value of $[HCO_3^-]$ indicates a primary metabolic alkalosis or a metabolic compensation for a primary respiratory acidosis while a low $[HCO_3^-]$ indicates a primary metabolic acidosis or a metabolic compensation for a primary respiratory acidosis. However, $[HCO_3^-]$ cannot be used as a true metabolic parameter, since changes in $PaCO_2$ also effect $[HCO_3^-]$.

The concept of the North American view is that *in vivo* data is used to calculate the expected rise or fall in $[HCO_3^-]$ and/or $PaCO_2$ that occur in specific acid-base disorders. The empirically derived *in vivo* information has been compiled from a large number of clinical studies in which the normal

compensatory reactions to each of the primary acid-base disorders has been investigated and quantified [23-30]. An observed value of $[HCO_3^-]$ or $PaCO_2$ below or above the expected value of $[HCO_3^-]$ or $PaCO_2$ is an indication for the presence and nature of a metabolic component or respiratory component of an acid-base disorder. Table 1 –1 presents the empirically found expected compensatory rise and fall in $[HCO_3^-]$ and $PaCO_2$ for the primary acid-base disturbances.

Table 1–1. Compensations to primary acid-base disturbances in the North American view [13].

disorder	primary change	compensatory response
metabolic acidosis	$\downarrow [HCO_3^-]$	1.2 mmHg decrease in $PaCO_2$ for every 1 mmol/l fall in $[HCO_3^-]$
metabolic alkalosis	$\uparrow [HCO_3^-]$	0.7 mmHg elevation in $PaCO_2$ for every 1 mmol/l rise in $[HCO_3^-]$
respiratory acidosis	$\uparrow PaCO_2$	
acute		1 mmol/l elevation in $[HCO_3^-]$ for every 10 mmHg rise in $PaCO_2$
chronic		3.5 mmol/l elevation in $[HCO_3^-]$ for every 10 mmHg rise in $PaCO_2$
respiratory alkalosis	$\downarrow PaCO_2$	
acute		2 mmol/l decrease in $[HCO_3^-]$ for every 10 mmHg fall in $PaCO_2$
chronic		5 mmol/l decrease in $[HCO_3^-]$ for every 10 mmHg fall in $PaCO_2$

1.4.3 The Scandinavian view; standard bicarbonate and base excess

The North American view requires calculations to be performed at the bedside of a patient. Moreover, to predict the amount of rise or fall in primary acid-base values, the acid-base disturbance of a patient should be known *a priori*. To overcome the 'problems' of bedside calculations and the paradox of classifying an already known acid-base disturbance, Astrup and Siggaard-Andersen developed the concept of the standard bicarbonate and the base excess as true metabolic acid-base parameters [11].

In 1960, Astrup described his equilibration method for the rapid measurement and calculation of the primary acid-base parameters pH, $PaCO_2$ and $[HCO_3^-]$ [12 , 31]. In a microtonometer a blood sample is equilibrated with two known CO_2 gas mixtures, one with a high $PaCO_2$ and one with a low $PaCO_2$. Plotting $PaCO_2$ and measured pH at both $PaCO_2$ values in a log $PaCO_2$–pH diagram, and connecting the two points with a line yields the *in vitro* CO_2 equilibration curve. By measuring pH of the original blood sample and putting it in the plot, the actual $PaCO_2$ of the blood sample can be read from the CO_2 equilibration curve. With the Henderson-Hasselbalch equation $[HCO_3^-]$ can be calculated.

With the log $PaCO_2$–pH chart and the *in vitro* CO_2 equilibration curve of a patient, $[HCO_3^-]$ can be calculated at any desired $PaCO_2$ value. Astrup proposed to use the $[HCO_3^-]$ of a blood sample at a $PaCO_2$ of 40 mmHg as a true metabolic parameter, since this would be the concentration that would have been found in the blood sample if the influence of the respiration was eliminated. He called it the standard bicarbonate concentration or SB.

At the same time, Siggaard-Andersen completed his titration experiments in which he determined the CO_2 equilibration curves of normal blood and blood with known amounts of non-volatile acids and bases at a fixed $PaCO_2$ of 40 mmHg. Based on these experiments he added to the log $PaCO_2$–pH diagram of Astrup a curved line representing the amount of non-volatile acid or base needed to titrate the blood sample at a $PaCO_2$ of 40 mmHg to a pH of 7.40 at a temperature of 37 °C. Astrup and Siggaard-Andersen called this the base excess or BE. Positive base excess values indicate a relative deficit of non-volatile acids while negative base excess values indicate a relative surplus of non-volatile acids. A base excess of 0 means that there is no

metabolic component in the acid-base disorder. In modern analysers, BE is calculated from pH, $PaCO_2$, $[HCO_3^-]$ and the haemoglobin concentration of the arterial blood sample at hand.

The most important argument against the use of standard bicarbonate and base excess is that they are determined *in vitro*. The *in vitro* CO_2 equilibration curve is the equilibration curve of whole blood in a tube or syringe. It has been shown that *in vivo* buffering of protons is different from the in *vitro* buffering of protons [17]. This is mainly because *in vivo* buffering takes place in the extracellular fluid in which the haemoglobin concentration (a powerful chemical buffer) is lower than in whole blood. Both Siggaard-Andersen himself and Severinghaus proposed to calculate BE not with the measured haemoglobin concentration of the sample, but with a haemoglobin concentration of 5 g/dl, which is the Hb concentration relative to the total volume of extracellular fluid of the body [32, 33]. This BE is also known as BEecf (Base Excess of extracellular fluid), SBE (Standard Base Excess) and BE5 (Base Excess at a haemoglobin concentration of 5 g/dl.

With BE as a true metabolic parameter, classifying acid-base disturbances is now straightforward. Figure 1 –1 and Table 1 –2 show all possible acid-base classifications based on pH, $PaCO_2$ and BE.

Table 1–2. Classification of acid-base disorders in the Scandinavian view. The signs '-', '+' and '=' indicate an observed value being respectively below, above or within its 95% normal reference interval. See also Error: Reference source not found.

	pH	$PaCO_2$	BE	classification
1	-	+	=	respiratory acidosis
2	-	+	+	partly compensated respiratory acidosis
3	=	+	+	compensated respiratory acidosis OR compensated metabolic alkalosis OR combined respiratory acidosis and metabolic alkalosis
4	+	+	+	partly compensated metabolic alkalosis
5	+	=	+	metabolic alkalosis
6	+	-	+	combined respiratory and metabolic alkalosis
7	+	-	=	respiratory alkalosis
8	+	-	-	partly compensated respiratory alkalosis

	pH	$PaCO_2$	BE	classification
9	=	-	-	compensated respiratory alkalosis OR compensated metabolic acidosis OR combined respiratory alkalosis and metabolic acidosis
10	-	-	-	partly compensated metabolic acidosis
11	-	=	-	metabolic acidosis
12	-	+	-	combined respiratory and metabolic acidosis
	=	=	=	normal
x	unclassifiable			

To determine whether an observed value for an acid-base parameter is too low, normal or too high, standard univariate 95% reference intervals are used. Table 1 –3 presents the associated upper and lower cut-off values for the univariate 95% reference intervals of arterial pH, $PaCO_2$, BE and $[HCO_3^-]$.

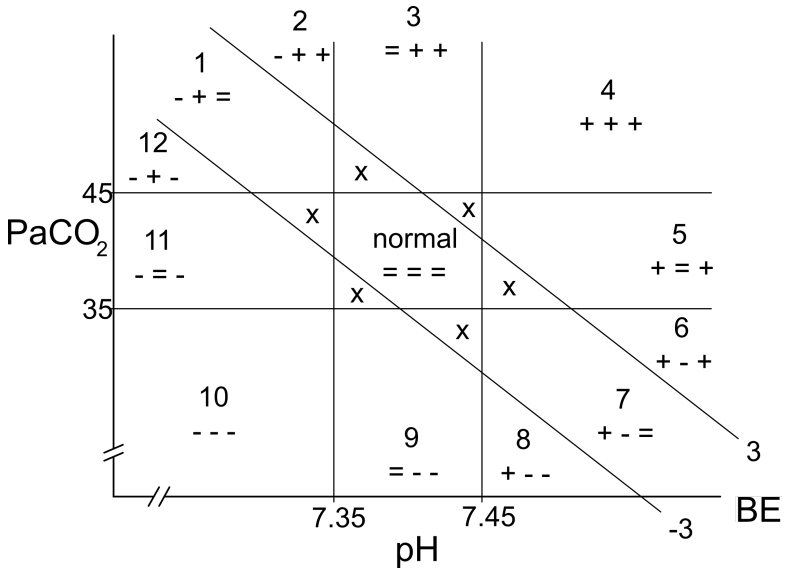

Figure 1–1. Areas of acid-base classification according to the method presented in Error: Reference source not found. Combinations of low, high or normal observed values yield 12 specific acid-base disorder regions. In the normal region, all values are within their standard 95% normal reference intervals. The 'x' regions are formally not classifiable [34]. In these regions, one of the three observed acid-base values is outside its 95% univariate reference interval.

Table 1–3. Upper and lower limits of the standard 95% normal reference intervals for the acid-base variables in arterial blood.

Acid-base variable	lower limit	upper limit
pH	7.35	7.45
$PaCO_2$	35 mmHg	45 mmHg
BE	-3 mmol/l	3 mmol/l
$[HCO_3^-]$	21 mmol/l	27 mmol/l

1.5 Objective and scope of this thesis

Two main problems are currently associated with the interpretation and evaluation of arterial acid-base measurements in an intensive care setting.

The first problem occurs when classifying the acid-base variables pH, $PaCO_2$ and BE according to the method described in section 1.4.3. A strict adherence to the classification rules as described in Table 1 –3 reveals that some combinations of observed values for the three acid-base variables can formally not be classified. This was found when an attempt was made to computerise the classification scheme of Astrup and Siggaard-Andersen in a rule-based expert system [34]. Typically, the *unclassifiable* situation occurs when only one of the three observed acid-base values is outside its 95% univariate reference interval, while the other two are within their 95% univariate intervals. In Figure 1 –1, this situation is represented by the triangular regions denoted by 'x'.

The second problem originates from the use of the 95% univariate reference interval as the standard statistical model for evaluating the 'normalcy' of observed arterial acid-base values from intensive care patients.

A first critical note on the use of 95% reference intervals is that the determination of the respective reference intervals and the characteristics of the reference population are completely unknown. In general, reference intervals are derived from a representative sample of a (often) 'healthy' reference population [35]. The process of defining the reference criteria, the selection of reference individuals, analytical considerations and the use of statistical techniques for defining valid 95% univariate reference intervals are described in detail [36-40]. Nothing is known, however, about the determination of the 95% univariate reference intervals that are presented in Table 1 –2. If we assume that the intervals are defined on a 'healthy' reference population, what is the value of these intervals in an intensive care setting where it is to be expected that most of the observed acid-base values will be outside these 'health'-based intervals?

A second critical note concerns the number of reference intervals used. Traditionally, the interpretation of the acid-base status involves the use of three separate 95% reference intervals for evaluating the acid-base variables: pH, $PaCO_2$ and $[HCO_3^-]$ in the North American view, or pH, $PaCO_2$ and BE in the Scandinavian view. From the Henderson-Hasselbalch equation (Equa-

tion 1 –4), however, it is apparent that the relationship between pH, log PCO_2 and log $[HCO_3^-]$ is a linear one. This can best be appreciated when Equation 1 –4 is rewritten as:

$$pH - \log [HCO_3^-] + \log PCO_2 = pK' - \log \alpha$$

(1–5)

with pK' and $\log \alpha$ both being constant.

Moreover, in Chapter Error: Reference source not found it will be demonstrated that the relationship between pH, $PaCO_2$ and BE is also (almost) linear. Consequently, as Madias [41] already pointed out, it is illogical and fundamentally wrong that three separate 95% univariate reference intervals are used, while only two of the three variables can change independently.

A third critical note on the use of univariate 95% reference intervals is that the 95% univariate interval is not the proper statistical model for evaluating arterial acid-base values. Theoretically, the use of more than one 95% univariate reference interval in case of a simultaneous evaluation of multiple variables – which is the case when interpreting arterial acid-base values – is prone to error and leads *a priori* to more false positive and false negative observations [42-44]. This will be illustrated in detail in Chapter Error: Reference source not found.

This thesis describes a new multivariate statistical reference model for evaluating and classifying arterial acid-base variables in an intensive care environment that addresses all of the above mentioned problems. The essence of the model is that a single 95% multivariate statistical reference region is defined on a large reference population consisting of acid-base data coming from intensive care patients themselves. Furthermore, the multivariate reference model is not defined on the original acid-base measurements but rather on the values obtained after applying a mathematical data reduction transformation procedure. Finally, based on the outcome of this transformation, a new way of classifying pH, $PaCO_2$ and BE values will be proposed that will have no *unclassifiable* categories, unlike the method described in 1.4.3.

The outline of this thesis is as follows. In Chapter Error: Reference source not found, the mathematical data reduction technique will be introduced, together with the results of various transformed large acid-base data sets coming from several ICUs. In Chapter Error: Reference source not found,

a two-dimensional graphical representation of the three acid-base variables will be presented, based on the mathematical transformation as described in Chapter Error: Reference source not found. Also, the new classification model for pH, $PaCO_2$ and BE combinations will be described. Then, in Chapter Error: Reference source not found, the technique for defining a 95% multivariate patient-based reference region for the acid-base variables will be described. Chapter Error: Reference source not found presents the computational methods involved in the data reduction transformation procedure and the construction of the multivariate reference model. It also presents the prototype computer programs that were built for defining multivariate acid-base reference regions and describes their use in daily clinical practice. Chapter Error: Reference source not found exemplifies the use and practicability of the proposed graphical representation of acid-base data using measurements from three intensive care patients. In Chapters Error: Reference source not found and Error: Reference source not found, the results of the clinical evaluation of the multivariate acid-base reference regions and classification model can be found. The thesis is concluded with a general discussion.

1.6 References

1. Solberg HE. Establishment and use of reference values. In: Burtis CA, Ashwood ER, eds. Tietz textbook of clinical chemistry. 2nd Edition, W.B. Saunders Company, 1994; 454-484.2.

2. Guder WG, Büttner J. Clinical chemistry in laboratory medicine in Europe-past, present and future challenges. *Eur J Clin Chem Clin Biochem* 1997; 35:487-494.3.

3. Sunderman FW, Sr. The foundation of clinical chemistry in the United States. *Clin Chem* 1994; 40:835-842.4.

4. Büttner J. Clinical chemistry as scientific discipline: historical perspectives. *Clin Chim Acta* 1994; 232:1-9.5.

5. Wold S. Chemometrics, why, what and where to next? *J Pharm Biomed Anal* 1991; 9:589-596.6.

6. Karjalainen EJ. The role of chemometrics in medical decision making. *Scand J Clin Lab Invest Suppl* 1990; 202:109-111.7.

7. Weil MH, Von Planta M, Rackow EC. Critical care medicine: introduction and historical perspective. In: Shoemaker W, ed. Textbook of Critical

Care. 2nd Edition, Philadelphia: W.B. Saunders Company, 1989; 1-5.8.
8. Muakkassa FF, Rutledge R, Fakhry SM, et al. ABG's and arterial lines: the relationship to unnecessarily drawn arterial gas samples. *J Trauma* 1990; 30:1087-1095.9.
9. Severinghaus JW, Astrup PB. History of blood gas analysis. Boston: Little, Brown and Company, 1987, Lange BP, ed., *International Anesthesiology Clinics*; vol 25.10.
10. Astrup P, Severinghaus JW. The history of blood gases, acids and bases. The history of blood gases, acids and bases. 1st Edition, Copenhagen: Munksgaard International Publishers, 1986; 264-295.11.
11. Astrup P, Jörgensen K, Siggaard-Andersen O, et al. The acid-base metabolism, a new approach. *Lancet* 1960;1035-1039.12. 12. Astrup P. A new approach to acid-base metabolism. *Clin Chem* 1961; 7:1-15.13.
13. Rose BD. Clinical Physiology of Acid-Base and Electrolyte Disorders. 4th Edition, New York: McGraw-Hill, Inc., 1994; 853.14.
14. Hainsworth R. Acid-base balance. Physiological Society Study Guides. Manchester: Manchester University Press, 1986; 155; vol 1.15.
15. Lane EE, Walker JF. Clinical arterial blood gas analysis. St. Louis: The C.V. Mosby Company, 1987; 247.16.
16. Bunker JP. The great trans-atlantic acid-base debate. *J Anesthesiol* 1965; 26:591-594.17.
17. Schwartz WB, Relman AS. A critique of the parameters used in the evaluation of acid-base disorders. "Whole-blood buffer base" and "standard bicarbonate" compared with blood pH and plasma bicarbonate concentration. *New Engl J Med* 1963; 268:1382-1388.18.
18. Rispens P, Zijlstra WG, Van Kampen EJ. Significance of bicarbonate for the evaluation of non-respiratory disturbances of acid-base balance. *Clin Chim Acta* 1974; 54:335-347.19.
19. Siggaard-Andersen O, Fogh-Andersen N. Base excess or buffer base (strong ion difference) as measure of a non-respiratory acid-base disturbance. *Acta Anaesthesiol Scand Suppl* 1995; 107:123-128.20.
20. Severinghaus JW. Acid-base balance nomogram. A Boston-Copenhagen détente. *Anesthesiology* 1976; 45:3-5.21.
21. Severinghaus JW. Acid-base balance controversy. Editorial introduction. *J Clin Monit* 1991; 7:274-275.22.
22. Severinghaus JW. Siggaard-Andersen and the "Great Trans-Atlantic Acid-Base Debate". *Scand J Clin Lab Invest* 1993; 53:99-104.23.
23. Arbus GS, Herbert LA, Levesque PR, et al. Characterization and clinical

application of the "significance band" for acute respiratory alkalosis. *New Engl J Med* 1969; 280:117-123.24.
24. Bushinsky DA, Coe FL, Katzenberg C, et al. Arterial PCO2 in chronic metabolic acidosis. *Kidney Int* 1982; 22:311-314.25.
25. Javaheri S, Shore NS, Rose B, et al. Compensatory hypoventilation in metabolic alkalosis. *Chest* 1982; 81:296-301.26.
26. Javaheri S, Kazemi H. Metabolic alkalosis and hypoventilation in humans. *Am Rev Respir Dis* 1987; 136:1011-1016.27.
27. Pierce NF, Fedson DS, Brigham KL, et al. The ventilatory response to acute base deficit in humans. Time course during development and correction of metabolic acidosis. *Ann Intern Med* 1970; 72:633-640.28.
28. Polak A, Haynie GD, Hays GM, et al. Effects of chronic hypercapnia on electrolyte and acid-base equilibrium. I. Adaptation. *J Clin Invest* 1961; 40:1223.29.
29. Van Yperselle de S, Brasseur L, De Coninck JD. The "carbon dioxide response curve" for chronic hypercapnia in man. *New Engl J Med* 1966; 275:117-122.30.
30. Gennari FJ, Goldstein MB, Schwartz WB. The nature of the renal adaptation to chronic hypocapnia. *J Clin Invest* 1972; 51:1722-1730.31.
31. Siggaard-Andersen O, Engel K, Jörgensen K, et al. A micro method for determination of pH, carbon dioxide tension, base excess and standard bicarbonate in capillary blood. *Scand J Clin Lab Invest* 1960; 12:172-176.32.
32.Siggaard-Andersen O. An acid-base chart for arterial blood with normal and pathophysiological reference areas. *Scand J Clin Lab Invest* 1971; 27:239-245.33.
33. Severinghaus JW. Acid-base balance controverse: case for standard-base excess as the measure of nonrespiratory acid-base imbalance. *J Clin Monit* 1991; 7:276-277.34. Wulkan RW.
34. Expert systems and multivariate analysis in clinical chemistry. Rotterdam: Erasmus University Rotterdam, 1992; 111 pp.35.
35. Solberg HE, Gräsback R. Reference values. *Adv Clin Chem* 1994; 27:1 -79.36.
36. Dybkær R, Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values, and International Committee for Standardization in Haematology (ICSH), Standing Committee on Reference Values. Approved Recommendation (1987) on the theory of reference values. Part 6. Presentation of observed values related to reference values. *J Clin Chem Clin*

Biochem 1987; 25:657-662.37.
37. PetitClerc C, Solberg HE. International Federation of Clinical Chemistry (IFCC). Approved Recommendation (1987) on the theory of reference values. Part 2. Selection of individuals for the production of reference values. *J Clin Chem Clin Biochem* 1987; 25:639-644.38.
38.Solberg HE, PetitClerc C. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values. Approved recommendation (1988) on the theory of reference values. Part 3. Preparation of individuals and collection of specimens for the production of reference values. *J Clin Chem Clin Biochem* 1988; 26:593-598.39.
39. Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values. Approved recommendation (1988) on the theory of reference values. Part 5. Statistical treatment of collected reference values. Determination of reference limits. *J Clin Chem Clin Biochem* 1987; 25:645-656.40.
40. Solberg HE. International Federation of Clinical Chemistry (IFCC), Scientific Committee, Clinical Section, Expert Panel on Theory of Reference Values, and International Committee for Standardization in Haematology (ICSH), Standing Committee on Reference Values. Approved Recommendation (1986) on the theory of reference values. Part 1. The concept of reference values. *J Clin Chem Clin Biochem* 1987; 25:337-342.41.
41. Madias NE, Adroqué HJ, Horowitz GL, et al. A redefinition of normal acid-base equilibrium in man: Carbon dioxide tension as a key determinant of normal plasma bicarbonate concentration. *Kidney Int* 1979; 16:612-618.42.
42. Stamhuis IH, Bezemer PD, Kuik D. Evaluation of univariate ranges with a multivariate standard. *J Clin Epidemiol* 1988; 41:359-366.43.
43. Solberg HE. Multivariate reference regions. *Scand J Clin Lab Invest Suppl* 1995; 222:3-5.44.
44. Schoen I, Brooks SH. Judgment based on 95% confidence limits. *Statistical Considerations* 1969; 53:190-193.

Chapter 2

The application of principal component analysis (PCA) to reduce the dimensionality of trivariate arterial acid-base data distributions

2.1 Introduction

In clinical practice, the interpretation of the arterial acid-base status is performed by a simultaneous evaluation of three acid-base laboratory parameters; pH of arterial blood, the partial pressure of carbon dioxide CO_2 in arterial blood ($PaCO_2$) for the evaluation of the respiratory component, and either the arterial bicarbonate-ion (HCO_3^-) concentration (a[HCO_3^-]) or the base excess (BE) for the evaluation of the metabolic component [1, 2]. Arterial acid-base values, however, are linearly related [1]. This means that if two of the three parameters are known, the third can be calculated. Thus, clinicians use three acid-base parameters to assess the acid-base status of a patient as if they were independent of each other, although only two of the three variables can change independently.

Although a strict linear relationship is not self-evident for pH, $PaCO_2$ and BE, an almost linear relationship is also present between these three variables. This was discovered during an earlier study [3] in which the distribu-

tions of large collections of pH, $PaCO_2$ and BE values were explored, using graphical software with capabilities of on-line three-dimensional rotation. During these explorations it was realised that, when plotting the combinations of the three variates as they occur in practice in three-dimensional space, the points are located on a surface with only a slight curvature.

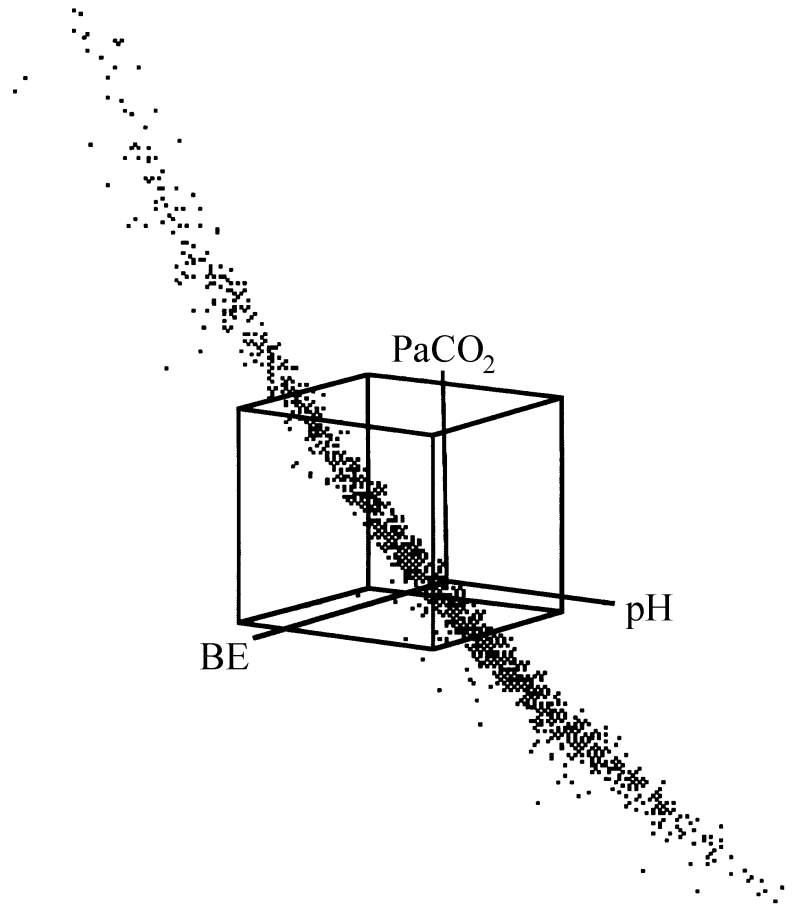

Figure 2–1. Three-dimensional view of a rotated pH, $PaCO_2$ and base excess (BE) data set. The cube represents the 95% reference volume as defined by the 95% univariate reference intervals of Error: Reference source not found.

Figure 2 –1 displays such a pH, $PaCO_2$ and BE distribution in three dimensions. The distribution is rotated in such a way that the curved plane of measurements can be easily viewed. The cube in the middle is the volume in three dimensions that represents the standard reference volume, as built from the three standard univariate 95% reference intervals for pH, $PaCO_2$ and BE of Table 2–1.

Table 2–1. Means (m) and standard deviations (s) as derived from the standard 95% reference intervals for the arterial acid-base variables. The standard deviations are calculated by assuming the 95% reference intervals to be 4 standard deviations wide.

	95 % reference interval	*m*	*s*
pH	7.35 – 7.45	7.40	0.025
$PaCO_2$ (mmHg)	35 – 45	40	2.5
$PaCO_2$ (kPa)	4.7 – 6.0	5.33	0.325
BE (mmol/l)	-3 – 3	0	1.5
$[HCO_3^-]$ (mmol/l)	21 – 27	24	1.5
log $PaCO_2$ (mmHg)	log 35 – log 40	log 40	(log 35 – log 40) / 4
log $PaCO_2$ (kPa)	log 4.7 – log 6.0	log 5.33	(log 4.7 – log 6.0) / 4
log $[HCO_3^-]$	log 21 – log 27	log 24	(log 21 – log 27) / 4

Having observed that the relationship between the arterial acid-base variables (both for combinations of pH, $PaCO_2$ and BE and pH, log $PaCO_2$ and log a$[HCO_3^-]$) is an almost linear one, the goal is to arrive at a mathematical description of this relationship and to investigate its departure from linearity. The mathematical technique to be used for such an investigation is a principal component analysis (PCA). PCA is a multivariate statistical technique for the compression of large data matrices [4, 5]. In this chapter, the results of a PCA of several distributions of acid-base data, coming from various ICUs, are described.

2.2 Materials and methods

2.2.1 Patient data

Six acid-base data sets from four different intensive care units were submitted to PCA. Data set *AZRbe* contains 1500 unselected combinations of pH, $PaCO_2$ and BE values from patients of the respiratory ICU of the Dijkzigt academic hospital, Rotterdam, The Netherlands. The term *unselected* means that no specific selection criteria were applied. In fact, all data sets are constructed by sampling the acid-base data as consecutively measured in the respective clinical laboratories. Data set *OLVGbe* contains 1500 un-

selected combinations of pH, $PaCO_2$ and BE values from patients of the general ICU of the OLVG hospital, Amsterdam, The Netherlands. Data set *OLVGab* comprises the 1500 combinations of pH, log $PaCO_2$ and log a[HCO_3^-] values from the same patients as the *OLVGbe* data set. Data set *SKZbe* contains 1500 combinations of pH, $PaCO_2$ and BE values from patients of the neonatal ICU of the Sophia Children's hospital, Rotterdam, The Netherlands. The data set is composed of equal numbers of data in three age groups: new-borns younger than five days, infants between five days and one month of age, and infants aged between one month and one year. Data set *ELIbe* contains 1500 unselected combinations of pH, $PaCO_2$ and BE values from the general ICU of the St. Elisabeth hospital, Tilburg, The Netherlands. Data set *ELIab* comprises the 1500 combinations of pH, log $PaCO_2$ and log a[HCO_3^-] from the same patients as the *ELIbe* data set.

2.2.2 Standardisation

Prior to the principal component analysis of an acid-base data set, each variable in the data set was standardised with fixed means and standard deviations according to:

$$z_i = \frac{x_i - m}{s}, i = 1, ..., N \tag{2–1}$$

where m and s are, respectively, the mean and standard deviation for the respective acid-base variables as presented in Table 2 –1, while N is the total number of cases in the data set. The z_i values are therefore the deviations from the mean m , measured in units of the corresponding standard deviation s.

2.2.3 Principal component analysis

The standardised data sets were then subjected to PCA. PCA is a mathematical transformation that enables the reduction of the number of variables in a multivariate data set whilst preserving as much of the original information as possible [4, 5]. Assuming a multivariate data set with p variables (x_1, x_2,..., x_p), PCA finds a new set of derived variables (z_1, z_2,...,z_p) that are

linear functions of x_1, x_2,..., x_p with the following properties:

- z_1 has maximum possible variance among all possible linear functions of x_1, x_2,..., x_p.
- z_k has maximum possible variance among all possible linear functions of x_1, x_2,...,x_p, subject to z_k being uncorrelated with z_1, z_2,...z_{k-1}, for $2 \leq k \leq p$ [4].

The derived variables z_1, z_2,...,z_p are called the principal components or PCs.

In linear algebraic terms, PCs are determined with an eigenvalue transformation of the variance-covariance matrix as derived from the multivariate data set. For a set of N vectors $\mathbf{x_i}$ $(i = 1, ..., N)$ in a p-dimensional data set, the variance-covariance matrix V is defined as:

$$V = \frac{\sum_{i=1}^{N} (\mathbf{x_i} - \mathbf{m})(\mathbf{x_i} - \mathbf{m})^{\mathbf{T}}}{N(N-1)}$$

(2–2)

where $\mathbf{m}$ is the vector of the mean of the set $\mathbf{x_i}$ $(i = 1, ..., N)$ and the superscript T indicates transposition of a vector, in the convention that an untransposed vector is a column vector. The eigenvalue transformation yields a transformation matrix U, which transforms the original vectors $\mathbf{m}$ into vectors $\mathbf{y}$, according to $\mathbf{y} = U\ \mathbf{x}$, such that the variance-covariance matrix $W = UVU^T$ of the transformed vectors $\mathbf{y}$ is a diagonal matrix. If U is constrained to be a unitary matrix, the component variances of the transformed vectors $\mathbf{y}$ appear as eigenvalues (λ) in the analysis and are found as the diagonal elements of W.

Since the eigenvalue transformation diagonalises the variance-covariance matrix, the total variance in the set of original vectors $\mathbf{x}$ is decomposed into p orthogonal directions. Thus, for a set of p-dimensional vectors $\mathbf{x}$ for which it is observed that most of the variance is confined to a subspace of dimension $l < p$, it is expected that the components 1 through l of the transformed vectors $\mathbf{y}$ contain most of the useful information. The components $l+1$ through p have only a small variance, and thus convey (almost) no information. In the present situation, $p = 3$ and due to the (almost) linear relationships between the variables in the standardised data sets, it is expected that $l = 2$.

2.3 Results

Table 2 –2 presents the results of the principal component analysis of each data set. The eigenvalues λ are shown for each of the three principal components (hereafter referred to as PC1, PC2 and PC3). The eigenvalues λ explain the contribution of each of the principal component to the total variance in the data set prior to PCA. For instance, in the *AZRbe* data set, PC1, PC2 and PC3 explain 62.37%, 36.91% and 0.71% of the total variance in the initial data set, respectively. From Table 2 –2 it can be concluded that for each data set, the percentage of variance explained by the third principal component (PC3) is only small compared to the variance explained by the first two principal components (PC1 and PC2) together. The explained variance by PC1 and PC2 for each data set is more than 99%. The data sets *OLVGab* and *ELIab* show the smallest explained variance by PC3; 0.03% and 0.09%, respectively. This is not surprising since these data sets consist of pH, log $PaCO_2$ and log a[HCO_3^-] values and these variables are linearly related according to the Henderson-Hasselbalch equation (see Chapter Error: Reference source not found) [1].

Table 2–2. Eigenvalues λ and contributions to the total variance in the initial data set (in brackets) for each principal component.

	PC1	PC2	PC3
AZRbe	20.54 (62.37%)	12.16 (36.91%)	0.235 (0.71%)
OLVGbe	26.04 (69.70%)	11.22 (30.03%)	0.101 (0.27%)
OLVGab	25.52 (74.30%)	8.82 (25.67%)	0.009 (0.03%)
SKZbe	21.09 (78.01%)	5.87 (21.72%)	0.072 (0.27%)
ELIbe	23.62 (64.55%)	12.83 (35.07%)	0.137 (0.37%)
ELIab	20.36 (58.91%)	14.17 (41.00%)	0.032 (0.09%)

For each data set, a matrix U can be built from the three separate normalised eigenvectors ε, which are used to calculate the associated principal component values from a combination of standardised original acid-base values. Table 2 –3 shows the normalised eigenvectors ε of each principal component for all data sets. With the matrices U, new trivariate distributions of principal component values were calculated from the original standardised acid-base data sets.

Table 2–3. Normalised eigenvectors ε of each principal component as obtained after PCA. The eigenvectors for PC1, PC2 and PC3 are columns 1, 2 and 3, respectively of the eigenmatrix matrix U. The eigenmatrix U will be the input of calculations to be presented in the next chapters.

	PC1	PC2	PC3
AZRbe	(0.297, -0.885, -0.358)	(0.709, -0.046, 0.703)	(0.639, 0.463, -0.614)
OLVGbe	(-0.039, -0.777, -0.628)	(0.742, -0.444, 0.503)	(0.669, 0.446, -0.594)
OLVGab	(0.044, 0.686, 0.727)	(-0.839, 0.420, -0.346)	(0.542, 0.594, -0.594)
SKZbe	(0.635, -0.757, 0.153)	(0.425, 0.508, 0.749)	(0.645, 0.411, -0.649)
ELIbe	(0.732, -0.184, 0.656)	(-0.184, 0.874, 0.450)	(0.656, 0.450, -0.605)
ELIab	(0.734, -0.012, 0.679)	(-0.391, 0.811, 0.436)	(0.556, 0.585, -0.590)

Table 2 –4 presents the characteristics of the resulting principal component value distributions. For each data set, the standard deviation of the PC3 distribution is small compared to the standard deviations of the PC1 and PC2 distributions. Data set *OLVGab* and *ELIab* have the smallest standard deviations for the third principal component value distribution: 0.097 and 0.178, respectively. This is in accordance with the results presented in Table 2 –2.

Table 2–4. Characteristics (m is mean and s is the standard deviation) of the principal component value distributions.

	PC1		PC2		PC3	
	m	*s*	*m*	*s*	*m*	*s*
AZRbe	1.282	4.532	0.695	3.487	0.177	0.485
OLVGbe	-0.773	5.103	-0.315	3.349	-0.215	0.317
OLVGab	0.357	5.052	0.250	2.969	-0.282	0.097
SKZbe	-3.358	4.593	-2.676	2.423	-0.109	0.268
ELIbe	-1.025	4.860	0.062	3.582	-0.005	0.370
ELIab	-1.008	4.513	-0.038	3.765	-0.071	0.178

Since for each data set the amount of explained variance is more than 99% when only PC1 and PC2 are considered, there is no significant loss of information when the acid-base values are projected onto the plane spanned by PC1 and PC2. Hence, any quantitative analysis based on PC1 and PC2 addresses the complete acid-base status. In Figure 2 –2, scatterplots of PC2 versus PC1 are shown for all data sets.

Since the plane of measurements in case of a pH, $PaCO_2$ and BE data set is slightly curved (see Figure 2 –1), it is interesting to investigate the effect of the curvature on the distribution of PC3 values. Therefore, the PC3 distribution characteristics of two data sets were investigated. This was done by constructing box-whisker plots of groups of PC3 values that are increasingly further away from the bivariate PC1-PC2 mean. As a cut-off point, a distance of 1 standard deviation score was chosen with a maximum of 10, yielding 11 groups of data. A box-whisker plot provides a graphical representation of the distribution of values in a given data set. The outer top and bottom horizontal lines of the box-whisker plots indicate the 95^{th} and 5^{th} percentiles of a distribution, respectively. The top and bottom horizontal

lines enclosing the box denote the 75^{th} and the 25^{th} percentile, respectively. The horizontal line inside the box denotes the median.

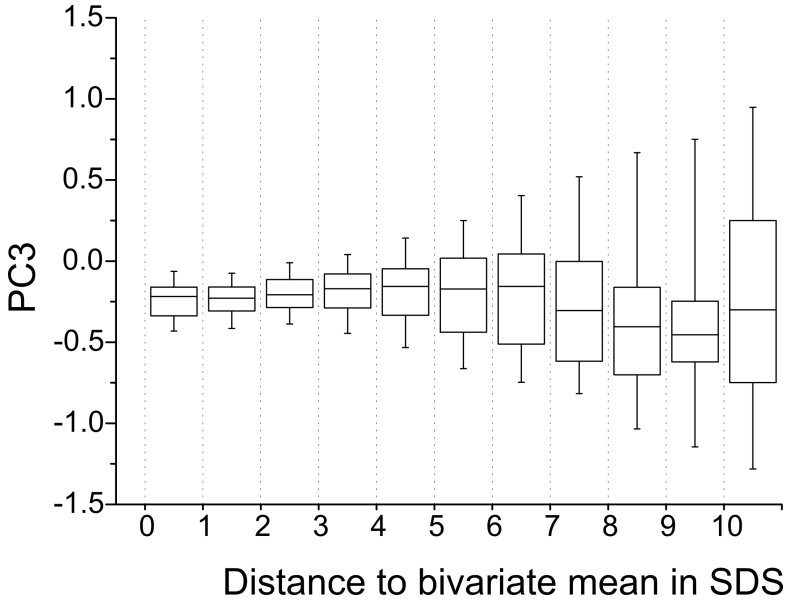

Figure 2–3. Box-whisker plots as a function of the distance from the mean in the PC1-PC2 plane for data set OLVGbe. SDS stands for 'standard deviation score'.

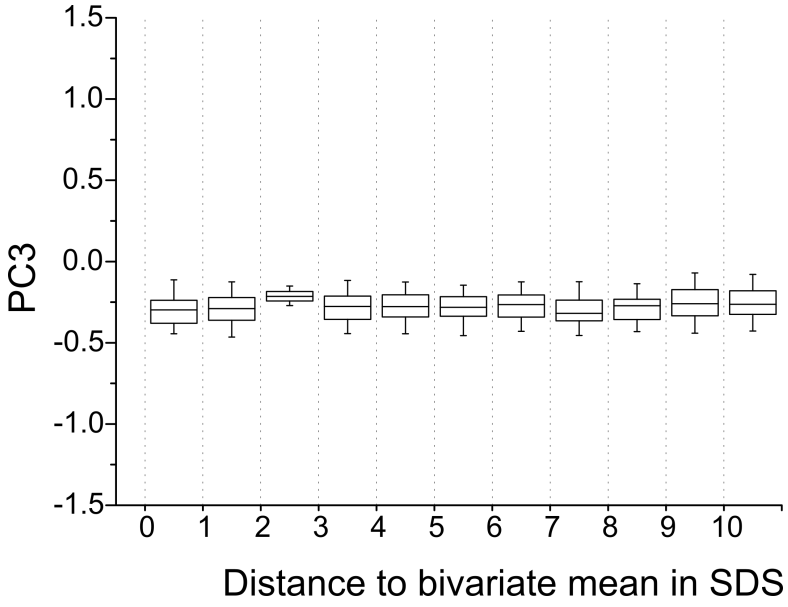

Figure 2–4. Box-whisker plots as a function of the distance from the mean in the PC1-PC2 plane for data set OLVGab.

Figure 2 –3 shows the box-whisker plots for the *OLVGbe* data set. In Figure 2 –4 the box-whisker plots are shown for the *OLVGab* data set. Comparing

both figures, it is apparent that with increasing distance from the mean in the plane spanned by the first two principal components PC1 and PC2, the variance in the PC3 distribution increases for data set *OLVGbe*, while the variance in the PC3 distribution of data set *OLVGab* remains the same for all distance strata. These figures illustrate the slight curvature of a PCA transformed pH, $PaCO_2$ and BE data set which is absent in a PCA transformed pH, log PaCO2 and log a$[HCO_3^-]$ dataset.

Figure 2 –5 presents a histogram of the 1500 calculated PC3 values of the transformed *OLVGab* distribution. The straight line in the normal probability plot in the upper part of Figure 2 –5 indicates that the 1500 PC3 values are normally distributed. This was confirmed with a Kolmogorov-Smirnov distribution fit test (D_{max} of 0.03 with a p-value of 0.118). Since these PC3 values are normally distributed, a parametric 95% reference interval may be derived from this distribution as $m \pm 2s$, resulting in a reference range of -0.472 to -0.092. The calculated PC3 value of a pH, log $PaCO_2$, log a$[HCO_3^-]$ combination from an ICU patient of the OLVG hospital, transformed with the corresponding eigenvectors of Table 2 –3, will have a probability of 95% of being located within this interval. A similar analysis, however, on the 1500 PC3 values of the transformed data set *ELIab* showed a bimodal distribution of PC3 values (Figure 2 –6). Consequently, the distribution was found to be significantly deviating from a normal distribution (D_{max} of 0.263 with a p-value < 0.01).

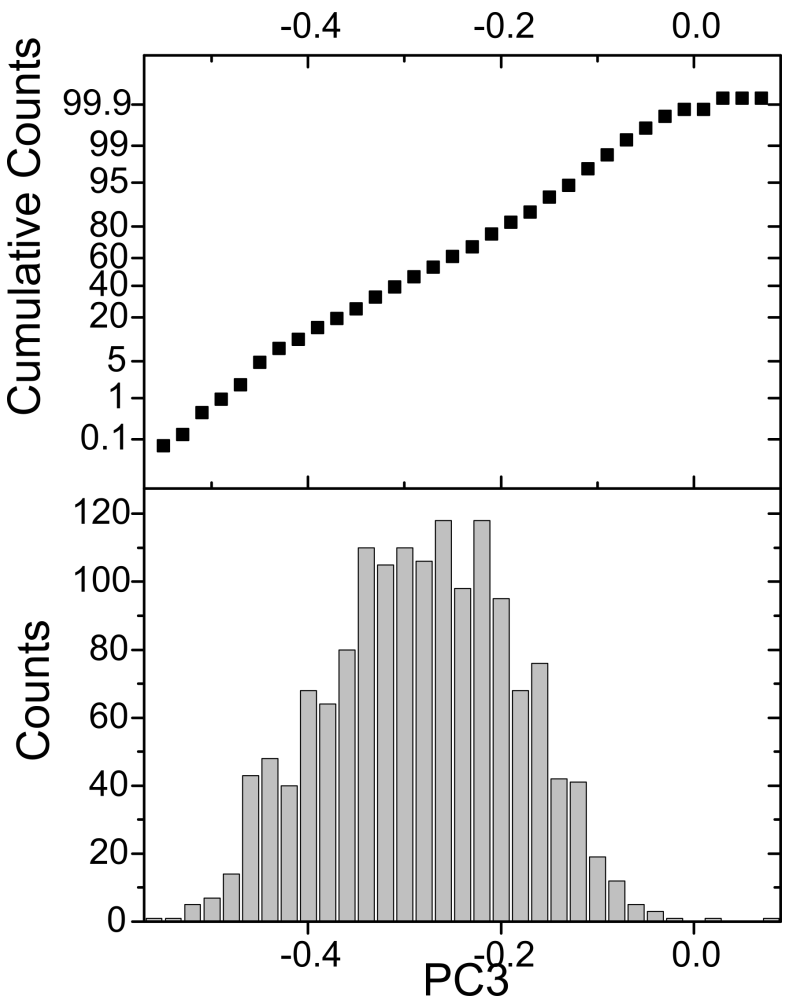

Figure 2–5. Histogram and normal probability plot of the 1500 third principal component values (PC3) of the OLVGab data set.

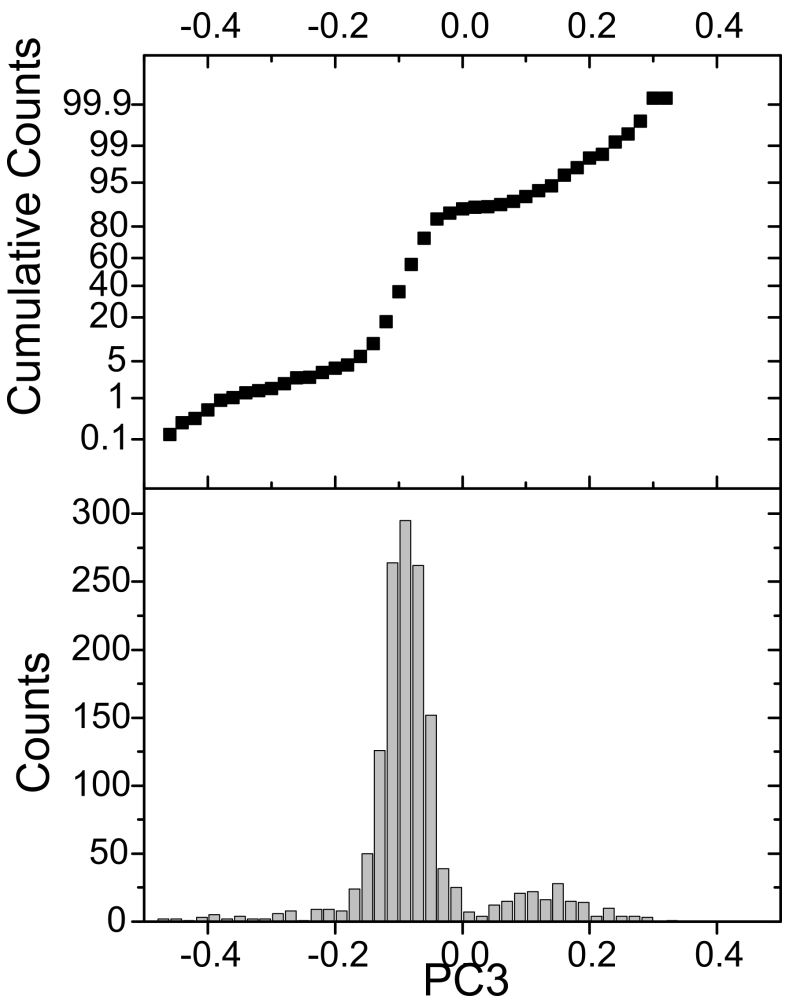

Figure 2–6. Histogram and normal probability plot of the 1500 third principal component values (PC3) of the ELIab data set.

2.4 Discussion

In 1979, Madias *et al.* [6] already noted that, when evaluating an acid-base status, it is illogical and fundamentally wrong to use pH, $PaCO_2$ as well as a$[HCO_3^-]$, while only two of these three variables are free to change independently. He proposed to evaluate acid-base disorders with only two of the three basic acid-base variables. This means, however, that clinicians are deprived of one of the three variables on which the interpretation of the acid-base status is traditionally based. In this chapter, a solution is proposed that allows a quantitative analysis of all three basic acid-base variables while being faithful to the interdependence between them. A multivariate statistical technique called principal component analysis (PCA) was used to

reduce the dimensionality of large trivariate distributions of acid-base variables. Results show that the acid-base status can be faithfully represented in a principal component subspace defined by the principal components PC1 and PC2, without significant loss of information. The distortion, measured as a percentage of variance not represented, is shown to be less than 0.7% for all the data sets investigated. The (small) percentage of explained variance by PC3 in data sets of pH, log $PaCO_2$ and log a$[HCO_3^-]$ (data sets *OLVGab* and *ELIab*) may be attributed to rounding effects and analytical imprecision. For the other data sets, consisting of pH, $PaCO_2$ and BE values, the curvature of the plane of measurements is an extra source of variance resulting in larger percentages of variance explained by PC3. However, this source of variance is only minor and for each data set it is therefore justified that quantitative analyses of acid-base disorders be based on PC1 and PC2 values after a PCA transformation, rather than on the original acid-base values. Furthermore, projection of the original points onto the PC1-PC2 subspace is (almost) distortionless. In Chapter Error: Reference source not found, this characteristic is used to define a sound way to graphically represent all three acid-base variables in a single two-dimensional representation.

The minor variance in PC3 may also serve as a plausibility check for acid-base laboratory values; each transformed combination of pH, $PaCO_2$ and a$[HCO_3^-]$/BE must lead to a small PC3 value. For a pH, log $PaCO_2$ and log a$[HCO_3^-]$ data set, PC3 must be within the 95% reference interval for PC3 as obtained from the PC3 values after PCA of an acid-base data set. For instance, the 95% reference interval for PC3 of the *OLVGab* data set was found to be -0.472 to 0.092. If a transformed combination of acid-base measurements is not within the interval, then it may be concluded that this specific combination of pH, $PaCO_2$ and a$[HCO_3^-]$ is not valid. Note that the interval is not equally centred around zero. From the definition of PCA one would expect that, when calculated means and standard deviations are used, the mean value for all principal component values would be zero. However, for the standardisation procedure the fixed means and standard deviations of Table 2 –1 were used, leading to the observation that the mean values of the principal components are different from 0 for the various data sets, since they have different means and variances for the original acid-base values.

Checking whether the PC3 value of a transformed acid-base observation is within the 95% reference interval is only possible for data sets of pH, log

$PaCO_2$ and log a[HCO_3^-], since the variance in PC3 is independent of the distance of an observation to the PC1-PC2 bivariate mean (see Figure 2 –6). Observations in a data set of pH, $PaCO_2$ and BE are located on a slightly curved plane of measurements, resulting in the effect that with increasing distances from the PC1-PC2 bivariate mean, the variance in PC3 increases (see Figure 2 –3). To check the plausibility of a transformed pH, $PaCO_2$ and BE combination one could either use the variance in PC3 as found for data with distances larger than or equal to 10 standard deviations scores (≥ 10), or use the variance in PC3 in the associated distance group.

One could argue that the relationship between the acid-base variables could be described by studying the formula used in acid-base analysers to calculate a[HCO_3^-] or BE from measured pH, $PaCO_2$ and haemoglobin. The advantage of the approach presented in this chapter, however, is that no prior knowledge is needed about the formula with which the a[HCO_3^-] or the BE are calculated. The method, therefore, adapts itself to the instruments used.

2.5 Acknowledgements

I am indebted to dr. R.N.M. Weijers and dr. D. Zandstra for making the data sets *OLVGbe* and *OLVGab* available. Dr. B. van der Berg and dr. R.W. Wulkan contributed similar material: the data sets *AZRbe* and *SKZbe*, respectively. Finally, dr. J.E. van Puyenbroek and dr. B. Speelberg provided the data sets *ELIbe* and *ELIab*.

2.6 References

1. Rose BD. Clinical Physiology of Acid-Base and Electrolyte Disorders. 4th Edition, New York: McGraw-Hill, Inc., 1994; 853.2.
2. Astrup P, Jörgensen K, Siggaard-Andersen O, et al. The acid-base metabolism, a new approach. *Lancet* 1960;1035-1039.3.
3. Gelsema ES, Leijnse B, Wulkan RW. A multi-dimensional analysis of three chemical quantities in the blood. *Med Inform* 1991; 16:43-54.4.
4. Jolliffe IT, Morgan BJT. Principal component analysis and exploratory factor analysis. *Stat Meth Med Res* 1992; 1:69-95.5.
5. Jolliffe IT. Principal Component Analysis. New York: Springer-Verlag,

1986, *Springer Series in Statistics*; vol 12.6.
6. Madias NE, Adroqué HJ, Horowitz GL, et al. A redefinition of normal acid-base equilibrium in man: Carbon dioxide tension as a key determinant of normal plasma bicarbonate concentration. *Kidney Int* 1979; 16:612-618.