CHAPTER 4

Material Balances and Data Consistency

Quantitative analysis of metabolism requires experimental data for the determination of metabolic fluxes, flux distributions, and measures of flux control (see Chapter 11), among other parameters. As such, these calculations exemplify methods and procedures for *upgrading the information content* of primary fermentation data. Whereas our focus in this book is on metabolism and its control, the basic philosophy of information content upgrade is applicable throughout the life sciences, so long as quantitative measurements are available.

Because these methods of information upgrade are data-driven, it is of the utmost importance to ensure the reliability of the data used. This can be done by applying the usual methods of random error minimization, *i.e.* use of repeat experiments, multiple sensors, careful calibrations, etc. An additional consideration (which is the subject of this chapter) is the introduction of *data redundancy* for the validation of both the actual measurements, and the broader mechanistic framework within which such measurements are

analyzed. For example, in the context of metabolic analysis, flux calculations are based on the measurement of the specific rates for substrate uptake and product formation, which represent the fluxes in and out of the cells. Before any such derivative calculations are carried out, it is important that the consistency of the data be confirmed, for example, the closure of the carbon balance.

Data redundancy is introduced when multiple sensors are employed for the measurement of the same variable or when certain constraints must be satisfied by the measurements so obtained, such as closure of material balances. Obviously, the greater the redundancy, the higher the degree of confidence for the data and their derivative parameters. Furthermore, redundancies can be employed for the systematic detection of the source of gross measurement errors or the identification of a particular element of the framework (*i.e.* model) most likely responsible for any observed inconsistencies. We demonstrate these ideas in this chapter in the context of fluxes, metabolism, and material balances. For this purpose, experimental data that are to be used for quantitative analysis must be

- *Complete.* This does not mean that *all* substrates and metabolic products must be measured, but those present in sufficient amounts should be quantified to allow the validation of the carbon and nitrogen (and, in some cases, also sulfur and phosphorus) balances. This requirement necessitates the use of defined, minimal media and essentially eliminates complex media from systematic metabolic studies.
- As much as possible, *noise free.* As discussed in Section 3.3, specific rates are derived from measurements of the concentration profiles, which make rate calculations difficult if these data are noisy. An important aspect of quantitative analysis of cellular metabolism therefore is the development of reliable and accurate analytical techniques, generally computerized high-performance bioreactors, where the most important culture variables are monitored on-line.

There are two approaches in assessing the consistency of experimental data. The first is based on a very simple metabolic model, the so-called *black box model*, where all cellular reactions are lumped into a single one for the overall cell biomass growth, and the method basically consists of validating elemental balances. It is rather easy to apply as the only information needed is that of the elemental composition of the substrates, metabolic products, and biomass, together with a set of fluxes in and out of the cell. The second approach recognizes far more biochemical detail in the overall conversion of substrates into biomass and metabolic products. As such, it is mathematically more involved, but, of course, it provides a more realistic depiction of the actual degrees of freedom than a black box model. We develop such

metabolic models in connection with our discussion of metabolic flux analysis (in Section 8.3). Because our focus in this chapter is mostly on methodology development for consistency analysis, we have eliminated unnecessary complications due to metabolic complexity and instead use a black box model.

4.1. THE BLACK BOX MODEL

In the black box model, cell biomass is the black box exchanging material with the environment, as depicted in Fig. 4.1, and processing it through many cellular reactions lumped into one, that of biomass growth. The fluxes in and out of the black box are given by the specific rates (in grams or moles of the compound per gram or mole of biomass and unit time). These are the specific substrate uptake rates (elements of $\mathbf{r}_s$) and the specific product formation rates (elements of $\mathbf{r}_p$). Additionally, there is accumulation of biomass within the box, which is represented as a flux with the specific rate μ. Because all cellular reactions are lumped into one overall reaction, the stoichiometric coefficients in this overall reaction are given by the yield coefficients introduced in Section 3.4:

$$X + \sum_{i=1}^{M} Y_{xp_i} P_i - \sum_{i=1}^{N} Y_{xs_i} S_i = 0 \tag{4.1}$$

where the specific rate of biomass formation is used as reference. Because the stoichiometric coefficient for biomass is 1, the forward reaction rate is given by the specific growth rate of the biomass, which, together with the yield coefficients of eq. (4.1), completely specifies the system. In the application of

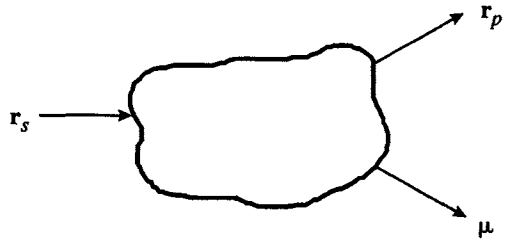

FIGURE 4.1 Representation of the black box model. The cell is considered as a black box, and fluxes in and out of the cell are the only variables measured. The fluxes of substrates into the cell are elements of the vector $\mathbf{r}_s$, and the fluxes of metabolic products out of the cell are elements of the vector $\mathbf{r}_p$. Some of the mass originally present in the substrates accumulates within the black box as formation of new biomass with the specific rate μ.

the black box model for analyzing data consistency, one may use either (1) a set of yield coefficients given in eq. (4.1) together with the specific growth rate; (2) a set of yield coefficients with respect to another reference, *e.g.*, one of the substrates, along with the specific rate of formation/consumption of this reference compound; (3) a set of specific rates for *all* substrates and products, including biomass; or 4) a set of all volumetric rates that are the product of the specific rates by the biomass concentration. All of these variables supply the same information. In the following, we will use either the yield coefficients given in eq. (4.1) or the set of specific rates that we collect in the total rate vector r, given by

$$\mathbf{r} = \begin{pmatrix} \mu \\ \mathbf{r}_p \\ -\mathbf{r}_s \end{pmatrix} = \begin{pmatrix} \mu & r_{p,1} & \cdots & -r_{s,1} & \cdots \end{pmatrix}^T \tag{4.2}$$

EXAMPLE 4.1

A Simple Black Box Model

Consider the aerobic cultivation of the yeast *Saccharomyces cerevisiae* on a defined, minimal medium, *i.e.*, glucose is the carbon and energy source and ammonia is the nitrogen source. During aerobic growth, the yeast oxidizes glucose completely to carbon dioxide. However, at very high glycolytic fluxes, a bottleneck in the oxidation of pyruvate leads to ethanol formation. Thus, at high glycolytic fluxes, both ethanol and carbon dioxide should be considered as metabolic products. Finally, water is formed in the cellular pathways, and this is also included as a product in the overall reaction. Thus, the black box model for this system is

$$X + Y_{xe}\ \text{ethanol} + Y_{xc}CO_2 + Y_{xw}H_2O - Y_{xs}\ \text{glucose} - Y_{xo}\ O_2 - Y_{xN}\ NH_3 = 0 \tag{1}$$

which can be represented with the specific rate vector:

$$\mathbf{r} = \begin{pmatrix} \mu & r_e & r_c & r_w & -r_s & -r_o & -r_N \end{pmatrix}^T \tag{2}$$

Obviously, the stoichiometric (or yield) coefficients of eq. (1) are not constant, as Y_{xe} is zero at low specific growth rates (corresponding to low glycolytic fluxes) and greater than zero for higher specific growth rates.

4.2. ELEMENTAL BALANCES

In the black box model, we have $M + N + 1$ variables: M yield coefficients for the metabolic products, N yield coefficients for the substrates, and the forward reaction rate μ [see eq. (4.1)] or the $M + N + 1$ specific rates of eq. (4.2). Because mass is conserved in the overall conversion of substrates to metabolic products and biomass, the $(M + N + 1)$ rates of the black box model are not completely independent but must satisfy several constraints. Thus, the elements flowing into the system must balance the elements flowing out of the system, *e.g.*, the carbon entering the system via the substrates has to be recovered in the metabolic products and biomass. Each element considered in the black box obviously yields one constraint. Thus, a carbon balance gives

$$1 + \sum_{i=1}^{M} h_{p,i} Y_{xp_i} - \sum_{i=1}^{N} h_{s,i} Y_{xs_i} = 0 \tag{4.3}$$

where $h_{s,i}$ and $h_{p,i}$ represent the carbon content (C-moles mole^{-1}) in the ith substrate and the ith metabolic product, respectively. In this equation, the elemental composition of biomass is normalized with respect to carbon, *i.e.*, it is represented by the form $CH_aO_bN_c$. The elemental composition of biomass depends on its macromolecular content and, therefore, on the growth conditions and the specific growth rate [*e.g.*, the nitrogen content is much lower under nitrogen-limited conditions than under carbon-limited conditions (see Table 4.1)]. However, except for extreme situations, it is reasonable to use the general composition formula $CH_{1.8}O_{0.5}N_{0.2}$ whenever the biomass composition is not known exactly.

The carbon balance of eq. (4.3) may also be formulated in terms of the specific rates. Thus, after multiplying eq. (4.3) by μ and applying the definition of the yield coefficients:

$$\mu + \sum_{i=1}^{M} h_{p,i} r_{p,i} - \sum_{i=1}^{N} h_{s,i} r_{s,i} = 0 \tag{4.4}$$

Often, we normalize the elemental composition of substrates and metabolic products with respect to their carbon content, *e.g.*, glucose is specified as CH_2O. Equation (4.3) is then written on per C-mole basis as

$$1 + \sum_{i=1}^{M} Y_{xp_i} - \sum_{i=1}^{N} Y_{xs_i} = 0 \tag{4.5}$$

TABLE 4.1 Elemental Composition of Biomass for Several Microorganisms[a]

Microorganism	Elemental composition	Ash content (w/w%)	Conditions
Candida utilis	$CH_{1.83}O_{0.46}N_{0.19}$	7.0	Glucose limited, D = 0.05 h^{-1}
	$CH_{1.87}O_{0.56}N_{0.20}$	7.0	Glucose limited, D = 0.45 h^{-1}
	$CH_{1.83}O_{0.54}N_{0.10}$	7.0	Ammonia limited, D = 0.05 h^{-1}
	$CH_{1.87}O_{0.56}N_{0.20}$	7.0	Ammonia limited, D = 0.45 h^{-1}
Klebsiella aerogenes	$CH_{1.75}O_{0.43}N_{0.22}$	3.6	Glycerol limited, D = 0.10 h^{-1}
	$CH_{1.73}O_{0.43}N_{0.24}$	3.6	Glycerol limited, D = 0.85 h^{-1}
	$CH_{1.75}O_{0.47}N_{0.17}$	3.6	Ammonia limited, D = 0.10 h^{-1}
	$CH_{1.73}O_{0.43}N_{0.24}$	3.6	Ammonia limited, D = 0.08 h^{-1}
Saccharomyces cerevisiae	$CH_{1.82}O_{0.58}N_{0.16}$	7.3	Glucose limited, D = 0.08 h^{-1}
	$CH_{1.78}O_{0.60}N_{0.19}$	9.7	Glucose limited, D = 0.255 h^{-1}
	$CH_{1.94}O_{0.52}N_{0.25}P_{0.025}$	5.5	Unlimited growth
Escherichia coli	$CH_{1.77}O_{0.49}N_{0.24}P_{0.017}$	5.5	Unlimited growth
	$CH_{1.83}O_{0.50}N_{0.22}P_{0.021}$	5.5	Unlimited growth
	$CH_{1.96}O_{0.55}N_{0.25}P_{0.022}$	5.5	Unlimited growth
	$CH_{1.93}O_{0.55}N_{0.25}P_{0.021}$	5.5	Unlimited growth
Pseudomonas fluorescens	$CH_{1.83}O_{0.55}N_{0.26}P_{0.024}$	5.5	Unlimited growth
Aerobacter aerogenes	$CH_{1.64}O_{0.52}N_{0.16}$	7.9	Unlimited growth
Penicillium chrysogenum	$CH_{1.70}O_{0.58}N_{0.15}$		Glucose limited, D = 0.038 h^{-1}
	$CH_{1.68}O_{0.53}N_{0.17}$		Glucose limited, D = 0.098 h^{-1}
Aspergillus niger	$CH_{1.72}O_{0.55}N_{0.17}$	7.5	Unlimited growth
Average	$CH_{1.81}O_{0.52}N_{0.21}$	6.0	

[a] The P content is given only for some microorganisms. The composition for *P. chrysogenum* is taken from Christensen *et al.* (1995), whereas the other data are taken from Roels (1983).

In eq. (4.5), the yield coefficients have the unit C-moles per C-mole biomass. Conversion to this unit from other units is illustrated in Box 4.1. Equation (4.5) (or eq. (4.4)) is very useful for checking the consistency of experimental data. Thus, if the sum of carbon in the biomass and the metabolic products does not equal the sum of carbon in the substrates, there is an inconsistency in the experimental data.

EXAMPLE 4.2

Carbon Balance in a Simple Black Box Model

We now return to the black box model of Example 4.1 and rewrite the conversion eq. (1) using the elemental composition of the substrates and

metabolic products specified. For biomass we use the elemental composition of $CH_{1.83}O_{0.56}N_{0.17}$, and therefore have

$$CH_{1.83}O_{0.56}N_{0.17} + Y_{xe}CH_3O_{0.5} + Y_{xc}CO_2 + Y_{xw}H_2O - Y_{xs}CH_2O - Y_{xo}O_2 - Y_{xN}NH_3 = 0 \quad (1)$$

Some may find it difficult to identify $CH_3O_{0.5}$ as ethanol, but the advantage of using the C-mole basis becomes apparent immediately when we look at the carbon balance:

$$1 + Y_{xe} + Y_{xc} - Y_{xs} = 0 \quad (2)$$

This simple equation is very useful for checking the consistency of experimental data. Thus, by using the classical data of von Meyenburg (1969), we find $Y_{xe} = 0.713$, $Y_{xc} = 1.313$, and $Y_{xs} = 3.636$ at a dilution rate of $D = 0.3$ h^{-1} in a glucose-limited continuous culture. Obviously the data are not consistent as the carbon balance does not close. A more elaborate data analysis (Nielsen and Villadsen, 1994) suggests that the missing carbon is ethanol, which could have evaporated as a result of ethanol stripping due to intensive aeration of the bioreactor.

By analogy to eq. (4.4), we find that a nitrogen balance gives

$$Y_{xN} = 0.17 \quad (3)$$

or in terms of specific rates:

$$r_N = 0.17\mu \quad (4)$$

If the measured rates of ammonia uptake and biomass formation do not conform with eq. (4), an inconsistency is identified in one of these two measurements or the nitrogen content of the biomass is different from that specified.

Similar to eq. (4.3), balances can be written for all other elements participating in the conversion [eq. (4.1)]. These balances can be conveniently written by collecting the elemental compositions of biomass, substrates, and metabolic products in the columns of a matrix E, where the first

BOX 4.1

Calculation of Yields with Respect to C-mole Basis

Yield coefficients are typically described as mol (g DW)$^{-1}$ or g (g DW)$^{-1}$. To convert the yield coefficients to a C-mol basis, information on the elemental composition and the ash content of biomass is needed. To illustrate the conversion, we calculate the yield of 0.5 g DW biomass (g glucose)$^{-1}$ on a C-mol basis. First, we convert the grams dry weight biomass to an ash-free basis, *i.e.*, determine the amount of biomass that is made up of carbon, nitrogen, oxygen, and hydrogen (and, in some cases, also phosphorus and sulfur). With an ash content of 8% we have 0.92 g ash-free biomass (g DW biomass)$^{-1}$, which gives a yield of 0.46 g ash-free biomass (g glucose)$^{-1}$. This yield can now be directly converted to a C-mol basis using the molecular masses in g C-mol^{-1} for ash-free biomass and glucose. With the standard elemental composition for biomass of $CH_{1.8}O_{0.5}N_{0.2}$, we have a molecular mass of 24.6 ash-free biomass C-mol^{-1}, and therefore we find a yield of 0.46/24.6 = 0.0187 C-mol biomass (g glucose)$^{-1}$. Finally, by multiplication with the molecular mass of glucose on a C-mol basis (30 g C-mol^{-1}), a yield of 0.56 C-mol biomass (C-mol glucose)$^{-1}$ is found.

column contains the elemental composition of biomass, columns 2 through $M + 1$ contain the elemental compositions of the M metabolic products, and columns $M + 2$ through $M + N + 1$ contain the elemental composition of the N substrates. If we consider I elements (normally four, namely, C, H, O and N), there are I rows in matrix E and the I elemental balances are represented by an equal number of algebraic equations similar to eq. (4.3), which can be summarized as

$$\mathbf{Er} = \mathbf{0} \tag{4.6}$$

With $N + M + 1$ specific rates (or volumetric rates) and I constraints, the degree of freedom is $F = M + N + 1 - I$. If exactly F rates are measured, it may be possible to calculate the other rates by using the I algebraic equations given by eq. (4.6), but, in this case, there are no redundancies left to check the consistency of the data. For this reason, it is advisable to strive for more measurements than the degrees of freedom of the system.

EXAMPLE 4.3

Elemental Balances in a Simple Black Box Model

We return to the black box model of Examples 4.1 and 4.2. With the elemental composition of biomass given previously, we can write the elemental composition matrix E as

$$\mathbf{E} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1.83 & 3 & 0 & 2 & 2 & 0 & 3 \\ 0.56 & 0.5 & 2 & 1 & 1 & 2 & 0 \\ 0.17 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{matrix} \leftarrow \text{carbon} \\ \leftarrow \text{hydrogen} \\ \leftarrow \text{oxygen} \\ \leftarrow \text{nitrogen} \end{matrix} \tag{1}$$

where the rows indicate the content of carbon, hydrogen, oxygen, and nitrogen, respectively, and the columns give the elemental composition of biomass, ethanol, carbon dioxide, water, glucose, oxygen, and ammonia, respectively. By using eq. (4.6), where **r** is replaced by a vector specifying the yield coefficients, we find

$$\begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 & 0 \\ 1.83 & 3 & 0 & 2 & 2 & 0 & 3 \\ 0.56 & 0.5 & 2 & 1 & 1 & 2 & 0 \\ 0.17 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ Y_{xe} \\ Y_{xc} \\ Y_{xw} \\ -Y_{xs} \\ -Y_{xo} \\ -Y_{xN} \end{pmatrix}$$

$$= \begin{pmatrix} 1 + Y_{xe} + Y_{xc} - Y_{xs} \\ 1.83 + 3Y_{xe} + 2Y_{xw} - 2Y_{xs} - 3Y_{xN} \\ 0.56 + 0.5Y_{xe} + 2Y_{xc} + Y_{xw} - Y_{xs} - 2Y_{xo} \\ 0.17 - Y_{xN} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix} \tag{2}$$

The first and last rows are identical to the balances derived in Example 4.2 for carbon and nitrogen, respectively. The balances for hydrogen and oxygen introduce two additional constraints. However, because the rate of water formation is impossible to measure, one of these equations must be used to calculate this rate (or yield). This leaves only one additional constraint from these two balances.

Equation (4.6) summarizes the balances for all elements. As discussed in Example 4.3, there is actually one fewer constraint as either the hydrogen or oxygen balance must be used to calculate the (nonmeasurable) water production rate. Obviously, water can be excluded by eliminating the yield coefficient for water between the O and H balances. A more elegant approach is to use the so-called *generalized degree of reduction* balance, which is derived as a linear combination of the elemental balances of eq. (4.6). This balance was introduced by Roels (1983) as a generalization of the earlier work of Erickson *et al.* (1978). It is generated by adding the elemental balances after multiplying them by a certain factor. By choosing appropriate multiplication factors, the yield coefficients (or rates) for *water, carbon dioxide*, and *nitrogen source* are eliminated from the resulting equation. To illustrate the procedure, consider the elemental balances given in Example 4.3, where we multiply the carbon balance by 4, the hydrogen balance by 1, the oxygen balance by -2, and the nitrogen balance by -3 to obtain the following:

4	$+4Y_{xe}$	$+4Y_{xc}$		$-4Y_{xs}$			$= 0$
1.83	$+3Y_{xe}$		$2Y_{xw}$	$-2Y_{xs}$		$-3Y_{xN}$	$= 0$
(-2) 0.56 +	(-2)0.5Y_{xe} +	(-2)2Y_{xc} +	(-2)Y_{xw}	-(-2)Y_{xs}	-(-2)2Y_{xo}		$= 0$
(-3) 0.17						-(-3)Y_{xN}	$= 0$
4.20	$+6Y_{xe}$			$-4Y_{xs}$	$+4Y_{xo}$		$= 0$

The resulting equation is the generalized degree of reduction balance for the system. Of course, this balance is not independent of the other elemental balances. It normally replaces either the oxygen or hydrogen balance, while the other is used to calculate the rate of water formation. Finally, one has the carbon, nitrogen, and degree of reduction balances to use for consistency analysis or for the calculation of unmeasured rates. In the original formulation of Erickson *et al.* (1978), the multiplication factor of each of the C, H, O, and N balances was interpreted as the number of free electrons available in C, H, O, and N, respectively, for transfer to oxygen upon the combustion of each element to *water, carbon dioxide*, and *ammonia* (as the nitrogen source). For the nitrogen balance the multiplier was always taken to be -3, as this is the predominant valency of nitrogen in biomass. In the generalized concept of Roels, the multiplication factors are arbitrary coefficients free to be chosen such that the resulting coefficients for water, carbon dioxide, and nitrogen source vanish. In this way, if another nitrogen source is used, *e.g.*, ammonium nitrate, a different multiplication factor is selected for the nitrogen balance in order to eliminate the yield coefficient for the nitrogen source from the generalized degree of reduction balance.

The coefficient multiplying the yield in the generalized degree of reduction balance is called the *degree of reduction* of the corresponding compound. For the preceding system, the degree of reduction is 4.2 for biomass, 6 for ethanol, 4 for glucose, 0 for water, ammonia, and carbon dioxide, and -4 for oxygen. With the generalization of Roels, the degree of reduction of the nitrogen-containing compounds depends on the nitrogen source used. In most cases ammonium is used as either the sole nitrogen source or in combination with another nitrogen source, yielding the following general expression for the degree of reduction κ of a compound with the elemental composition $CH_aO_bN_c$:

$$\kappa = 4 + a - 2b - 3c \tag{4.7}$$

Table 4.2 of the following section lists the degrees of reduction for compounds typically encountered in fermentation processes. Roels (1983), or the more recent publication of Nielsen and Villadsen (1994), can be consulted for further elaboration on the concept of the degree of reduction. With the introduction of a compound's degree of reduction κ, the generalized degree of reduction balance for any system is given by

$$\kappa_x + \sum_{i=1}^{M} \kappa_{p,i} Y_{xp_i} - \sum_{i=1}^{N} \kappa_{s,i} Y_{xs_i} = 0 \tag{4.8}$$

This balance is very useful as it is simple to set up and, together with the carbon and nitrogen balances, contains all the constraints imposed by the four elemental balances.

EXAMPLE 4.4

Analysis of Data Consistency in Anaerobic Yeast Cultivations

To illustrate the application of the generalized degree of reduction to the analysis of data consistency, we consider data from the anaerobic continuous cultures of *S. cerevisiae* obtained by Schulze (1995). Yield coefficients (all in C-moles or moles per C-mole biomass), obtained under conditions of glucose limitation, are listed for glucose (Y_{xs}), ethanol (Y_{xe}), carbon dioxide (Y_{xc}), and glycerol (Y_{xg}):

Dilution rate h^{-1}	Y_{xs}	Y_{xe}	Y_{xc}	Y_{xg}
0.1	7.81	3.88	2.13	0.67
0.2	8.06	4.00	2.26	0.73

First, the carbon balance

$$1 + Y_{xe} + Y_{xc} + Y_{xg} - Y_{xs} = 0 \tag{1}$$

is satisfied within 2% at $D = 0.1$ h^{-1} and within 1% at $D = 0.2$ h^{-1}. Such deviations (estimated relative to the carbon supply in the form of glucose) are very satisfactory.

The generalized degree of reduction balance gives

$$\kappa_x + 6Y_{xe} + 4.67Y_{xg} - 4Y_{xs} = 0 \tag{2}$$

The elemental composition of yeast was determined to be $CH_{1.78}O_{0.60}N_{0.19}$. Thus, the degree of reduction of biomass is 4.01, and, upon substitution of the yield coefficients from the preceding table, the generalized degree of reduction balance is found to close within 3% in both cases (again the deviation is given relative to glucose). It is interesting to note that for $D = 0.1$ h^{-1} the degree of reduction of the "missing carbon" is found (by invoking the carbon balance as well) to be very close to 6, indicating that the ethanol measurement may be underestimated. This is a general problem in yeast and other cultivations where volatile compounds are produced, and a loss of carbon in the neighborhood of less than 2% in these cultivations generally is considered acceptable.

4.3. HEAT BALANCE

In the conversion of substrates to metabolic products and biomass, part of the Gibbs free energy in the substrates is dissipated to the surrounding environment as heat. Especially under aerobic conditions, the energy dissipation may be substantial. Energy dissipation is determined by the difference between the total Gibbs free energy in the substrates and the total Gibbs free energy recovered in the metabolic products and biomass. The energy dissipation normally gives rise to changes in both the enthalpy and entropy of the system, and it is difficult to quantify (see also Section 13.1). Attention is, therefore, generally focused on heat production determined by the change in enthalpy, as this heat production has direct consequences for process cooling requirements for temperature control. With the black box model, the heat production Q_{heat} [kJ (C-mol biomass)$^{-1}$] of the overall process can be

calculated from

$$Q_{\text{heat}} = -\Delta H_c^0 = \sum_{i=1}^{N} Y_{xs_i} \Delta H_{c,i}^0 - \Delta H_{c,x}^0 - \sum_{i=1}^{M} Y_{xp_i} \Delta H_{c,i}^0 \quad (4.9)$$

where $\Delta H_{c,i}^0$ is the heat of combustion [kJ (C-mol)$^{-1}$] of the ith compound at standard conditions (298 K and 1 atm). The yield coefficients in the preceding equation are given on a C-mole basis. Table 4.2 lists the heats of combustion for some compounds typically found in fermentation media. Notice that Q_{heat} is actually a yield rather than a rate. To determine the rate of heat production, the preceding equation is multiplied by the growth rate. Equation (4.9) is useful for calculating the heat production from the yield coefficients and can be used for designing the cooling capacity of a bioreactor, as illustrated in Example 4.5.

EXAMPLE 4.5

Heat Generation at Anaerobic versus Aerobic Growth

We consider the growth of *S. cerevisiae* under anaerobic and aerobic conditions. The black box models for these two growth conditions can be taken to be

$$CH_{1.62}O_{0.53}N_{0.15} + 4.78CH_3O_{0.5} + 2.42CO_2 + 0.41H_2O - 8.20CH_2O - 0.15NH_3 = 0 \quad (1)$$

for anaerobic growth and

$$CH_{1.62}O_{0.53}N_{0.15} + 0.67CO_2 + 1.08H_2O - 1.67CH_2O - 0.15NH_3 - 0.64O_2 = 0 \quad (2)$$

for aerobic growth. We calculate the heat production for the two reactions to be

$$Q_{\text{anarob}} = [(8.20)(467) + (0.15)(383) - 560 - (4.78)(683)]$$
$$= 62.11 \text{ kJ (C-mol biomass)}^{-1} \quad (3)$$
$$Q_{\text{aerob}} = [(1.67)(467) + (0.15)(383) - 560]$$
$$= 277.3 \text{ kJ (C-mol biomass)}^{-1} \quad (4)$$

TABLE 4.2 Heats of Combustion for Various Compounds at Standard Conditions (298 K and 1 atm) and pH 7[a]

Compound	Formula	Degree of reduction	$\Delta H^o_{c,i}$ (kJ C-mol^{-1})
Acetaldehyde	C_2H_4O	5	583
Acetic acid	$C_2H_4O_2$	4	437
Acetone	C_3H_6O	5.33	597
Ammonia	NH_3		383[c]
Biomass	$CH_{1.8}O_{0.5}N_{0.2}$	4.2	560
n-Butanol	$C_4H_{10}O$	6	669
Butyric acid	$C_4H_8O_2$	5	546
Citric acid	$C_6H_8O_7$	3	327
Ethane	C_2H_6	7	780[c]
Ethanol	C_2H_6O	6	683
Formaldehyde	CH_2O	4	571[c]
Formic acid	CH_2O_2	2	255
Fructose	$C_6H_{12}O_6$	4	469
Fumaric acid	$C_4H_4O_4$	3	334
Galactose	$C_6H_{12}O_6$	4	468
Glucose	$C_6H_{12}O_6$	4	467
Glycerol	$C_3H_8O_3$	4.67	554
Isopropanol	C_3H_8O	6	673
Lactic acid	$C_3H_6O_3$	4	456
Lactose	$C_{12}H_{22}O_{11}$	4	471
Malic acid	$C_4H_6O_5$	3	332
Methane	CH_4	8	890[c]
Methanol	CH_4O	6	727
Oxalic acid	$C_2H_2O_4$	1	123
Palmitic acid	$C_{16}H_{32}O_2$	5.75	624[b]
Propane	C_3H_8	6.67	740[c]
Propionic acid	$C_3H_6O_2$	4.67	509
Succinic acid	$C_4H_6O_4$	3.5	373
Sucrose	$C_{12}H_{22}O_{11}$	4	470
Urea	CH_4ON_2		632
Valeric acid	$C_5H_{10}O_2$	5.2	568

[a] The heat of combustion is given with the reference being CO_2, H_2O and N_2.
[b] Solid form.
[c] Gaseous form.

We see that much more heat is generated in the aerobic process [corresponding to 165 kJ (C-mol glucose metabolized)$^{-1}$] than in the anaerobic process [corresponding to about 8 kJ (C-mol glucose)$^{-1}$]. Thus, in the aerobic process, a large fraction of the free energy originally present in glucose dissipates as heat, whereas in the anaerobic process it is retrieved in ethanol. To illustrate the cooling requirements of a large-scale bioreactor, we calculate the total heat production for a typical industrial baker's yeast fermentation. We use a bioreactor volume of 100 m^3 and a biomass concentration of 50 g L^{-1} (corresponding to about 1.96 C-mol L^{-1}). For the batch phase, of such a process the specific growth rate is approximately equal to 0.25 h^{-1}. Using these data we first find the specific rate of heat production:

$$r_q = Q_{aerob}\,\mu = \left[277.3 \text{ kJ (C-mol biomass)}^{-1}\right](0.25 \text{ h}^{-1})$$

$$= 69 \text{ kJ (C-mol biomass)}^{-1} \text{ h}^{-1} \qquad (5)$$

and from here we find the total heat production to be

$$(69 \text{ kJ (C-mol biomass)}^{-1} \text{ h}^{-1})(1.96 \text{ C-mol L}^{-1})(100.000 \text{ L}) = 3.8 \text{ MW} \qquad (6)$$

This large heat production clearly illustrates the requirement for large amounts of cooling water to maintain a constant temperature in the bioreactor.

If the heat production rate can be measured accurately, such as by using a calorimeter [as illustrated in several publications; see, for example Larsson *et al.* (1991) and von Stockar and Birou (1989)] or measuring the temperature change in the bioreactor, the heat balance [eq. (4.9)] may be used to supply an additional redundancy along with the elemental balances of Section 4.2. If, however, the heat production rate cannot be measured (as would be the case of an anaerobic process where the heat production is very small), the introduction of an additional equation does not change the degrees of freedom due to the additional unknown variable (Q_{heat}). For aerobic processes, it is generally found that the rate of heat production is proportional to the oxygen uptake rate:

$$Q_{heat} = aY_{xo} \qquad (4.10)$$

Equation (4.10) is empirically found to be valid for microbial growth on different substrates with a proportionality constant approximately equal to 460 kJ per mol O_2 [see Table 4.3 and Example 4.5, where it was found to be

TABLE 4.3 Comparison of Y_{xo} and Q_{heat} for Bacteria Grown on Different Carbon Sources[a]

Substrate	Y_{xo} [mmol O_2 (g DW)$^{-1}$]	Q_{heat} [kJ (g DW)$^{-1}$]	Q/Y_{xo} [kJ (mol O_2)$^{-1}$]
Malate	30.6	14.0	458
Acetate	44.6	19.9	446
Glucose	21.3	10.0	469
Methanol	71.0	34.9	492
Ethanol	51.2	23.2	453
Isopropanol	135.8	56.5	416
n-Paraffins	62.5	26.2	419
Methane	156.3	68.6	439

[a] The data are taken from Abbott and Clamen (1973).

equal to 433 kJ (mol O_2)$^{-1}$]. Equation (4.10) may also be derived from a generalized degree of reduction balance, where the reference for nitrogen-containing compounds is taken to be N_2 (Roels, 1983; Nielsen and Villadsen, 1994). A consequence of eq. (4.10) is that the measurement of the rate of heat production is well-suited for checking the measurements of the oxygen uptake rate or as an alternative to this measurement.

4.4. ANALYSIS OF OVERDETERMINED SYSTEMS—IDENTIFICATION OF GROSS MEASUREMENT ERRORS

If there are more measurements available than the degrees of freedom F, the system is generally called *overdetermined.* In this case the redundancy of the measurements can be used to (a) calculate the rates of nonmeasured metabolites; (b) increase the accuracy of the available measurements through the application of essentially a least squares calculation; and (c) identify the most likely source of gross measurement errors or even the source of inconsistencies in the formulation of the black box framework. This can be carried out in a straightforward manner. For example, if only one rate is not measured, we can use the carbon balance to calculate that rate and the remaining (nitrogen and degree of reduction) balances to check the overall consistency of the data. A more effective analysis is based on the *simultaneous* use of all balances, elemental and otherwise, for the calculation of the nonmeasured rates as well as for data consistency analysis. This is best carried out through the use of matrix manipulations. We follow this procedure here; however, we have reduced the use of matrix operations to a minimum in order to facilitate

the review of this material by those with limited exposure to this subject. Furthermore, we have provided a rudimentary review of matrix operations in Box 4.2. Finally, we provide several examples to illustrate these operations when applied to our system of aerobic yeast cultivation without ethanol formation.

We begin our analysis with the elemental balances of eq. (4.6), which we rewrite in the following form by partitioning the rate vector $\mathbf{r}$ into two vectors: One, $\mathbf{r}_m$, collects all measured rates, and another, $\mathbf{r}_c$, collects the remaining rates (that need to be calculated, hence the subscript c):

$$\mathbf{E}\mathbf{r} = \mathbf{E}_c\mathbf{r}_c + \mathbf{E}_m\mathbf{r}_m = 0 \tag{4.11}$$

Similarly, the elemental matrix $\mathbf{E}$ is partitioned by separating the columns with the elemental composition of the compounds that have been measured into one matrix $\mathbf{E}_m$ and the columns of the nonmeasured compounds (that must be calculated from the balances) into matrix $\mathbf{E}_c$. Of course, if exactly F variables are measured, there are just enough equations to determine the nonmeasured rates. In this case $\mathbf{E}_c$ is a square matrix with dimensions (I x I) equal to the number of constraints (or balances; I). If it has full rank, *i.e.*, $\text{rank}(\mathbf{E}_c) = I$ (see Box 4.2), the nonmeasured specific rates of $\mathbf{r}_c$ can be calculated by solving eq. (4.11):

$$\mathbf{r}_c = -\mathbf{E}_c^{-1}\mathbf{E}_m\mathbf{r}_m \tag{4.12}$$

If $\mathbf{E}_c$ is square and has full rank, the system is called *observable*, as there are exactly enough measurements to determine the unknown rates, *i.e.*, the system is overdetermined. If more rates are measured than the degrees of freedom F, there are more equations available than the minimum needed for the determination of the (now fewer) unknown rates. In this case, a least squares approach is usually employed, whereby the unknown rates are calculated from a combination of the available balances in order to increase the accuracy of the so-obtained estimates. The matrix equivalent of this situation is that the elemental submatrix $\mathbf{E}_c$ now is not square, and its inverse therefore cannot be determined. However, multiplication of eq. (4.11) by the transpose matrix $\mathbf{E}_c^T$ (see Box 4.2) of $\mathbf{E}_c$ yields

$$\mathbf{E}_c^T(\mathbf{E}_c\mathbf{r}_c + \mathbf{E}_m\mathbf{r}_m) = (\mathbf{E}_c^T\mathbf{E}_c)\mathbf{r}_c + \mathbf{E}_c^T\mathbf{E}_m\mathbf{r}_m = 0 \tag{4.13}$$

$\mathbf{E}_c^T\mathbf{E}_c$ is certainly square (see Box 4.2), and if it has full rank, it can be inverted to give the solution for $\mathbf{r}_c$:

$$\mathbf{r}_c = -\mathbf{E}_c^{\#}\mathbf{E}_m\mathbf{r}_m \tag{4.14}$$

BOX 4.2

Matrix Operations

A matrix is simply a set of numbers arranged in some array. The arrays that you will use in this text will usually be in the form of either a vector with multiple components or a square/rectangular matrix. In this box we give an introduction to the most simple matrix operations that are used throughout this text [for more details, see Strang (1988)].

Consider the generalized matrix A shown with two rows and two columns:

$$\mathbf{A}\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}$$

Note that in matrix notation $A_{i,j}$ refers to the element of the ith row and the jth column, where $i = 1..n$, and $j = 1..m$. The dimension of a matrix is specified as n x m, and in this case A is a 2 x 2 matrix.

Basic Matrix Operations

Consider matrix A already shown and another 2 x 2 matrix B:

$$\mathbf{B}\begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}$$

The sum and difference of matrices A and B give 2 x 2 matrixes C and D, respectively:

$$\mathbf{C} = \mathbf{A} + \mathbf{B} = \begin{pmatrix} A_{1,1} + B_{1,1} & A_{1,2} + B_{1,2} \\ A_{2,1} + B_{2,1} & A_{2,2} + B_{2,2} \end{pmatrix}$$

$$\mathbf{D} = \mathbf{A} - \mathbf{B} = \begin{pmatrix} A_{1,1} - B_{1,1} & A_{1,2} - B_{1,2} \\ A_{2,1} - B_{2,1} & A_{2,2} - B_{2,2} \end{pmatrix}$$

Matrices are multiplied by numerical constants one component at a time:

$$\mathbf{E} = 2\mathbf{A} = 2\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} = \begin{pmatrix} 2A_{1,1} & 2A_{1,2} \\ 2A_{2,1} & 2A_{2,2} \end{pmatrix}$$

The multiplication operation, of a matrix by a vector or a matrix by a matrix, is less apparent and is illustrated next. Consider a vector **v**, where **v** has the same number of elements as the number of columns of **A**:

$$\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$$

The product of **A** and **v** is then specified as

$$\mathbf{F} = \mathbf{Av} = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} A_{1,1}v_1 + A_{1,2}v_2 \\ A_{2,1}v_1 + A_{2,2}v_2 \end{pmatrix}$$

Note that in essence each element of **F** represents the sum of the products of the corresponding rows of **A** and **v**. For the general case of a matrix **A** having dimensions m x n, and a vector **v** with dimensions r x 1, it is obviously necessary that r must be equal to n; (*i.e.*, the number of columns of **A**), and the product matrix will have dimensions m x n.

In a similar fashion, one can multiply two matrices of compatible dimensions as shown next, where in this case each row of **A** is multiplied by the corresponding column of **B** to give the 2 x 2 matrix **G**:

$$\mathbf{G} = \mathbf{AB} = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}$$

$$= \begin{pmatrix} A_{1,1}B_{1,1} + A_{1,2}B_{2,1} & A_{1,1}B_{1,2} + A_{1,2}B_{2,2} \\ A_{2,1}B_{1,1} + A_{2,2}B_{2,1} & A_{2,1}B_{1,2} + A_{2,2}B_{2,2} \end{pmatrix}$$

Note that matrix multiplication is associative, i.e., **(AB)C = A(BC)**, and distributive, *i.e.*, **A(B + C) = AB + BC**, but *not* commutative, *i.e.*, **AB ≠ BA**.

(continues)

(continued)

Example 1

Consider matrices A and B and vector v with the following numerical values:

$$\mathbf{A} = \begin{pmatrix} 0 & 1 \\ 2 & 3 \end{pmatrix}, \quad \mathbf{B} = \begin{pmatrix} 4 & 3 \\ 2 & 2 \end{pmatrix}, \quad \mathbf{v} = \begin{pmatrix} 3 \\ 5 \end{pmatrix}$$

$$\mathbf{A} + \mathbf{B} = \begin{pmatrix} 4 & 4 \\ 4 & 5 \end{pmatrix} \quad \mathbf{A} - \mathbf{B} = \begin{pmatrix} -4 & -2 \\ 0 & 1 \end{pmatrix}$$

$$2\mathbf{A} = \begin{pmatrix} 0 & 2 \\ 4 & 6 \end{pmatrix} \quad \mathbf{Av} = \begin{pmatrix} 5 \\ 21 \end{pmatrix} \quad \mathbf{AB} = \begin{pmatrix} 2 & 2 \\ 14 & 12 \end{pmatrix}$$

The reader is strongly encouraged to reproduce these results as an exercise.

Matrix Transpose

The transpose of matrix A, denoted as $\mathbf{A}^T$, is a matrix whose columns are taken directly from the rows of A, *i.e.*, row i of A becomes column i of $\mathbf{A}^T$. Thus, for the general case:

$$\mathbf{A}^T = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}^T = \begin{pmatrix} A_{1,1} & A_{2,1} \\ A_{1,2} & A_{2,2} \end{pmatrix}$$

Note: The transpose of AB is $(\mathbf{AB})^T = \mathbf{A}^T\mathbf{B}^T$.

Matrix Inverse

The inverse of an n x n matrix A, denoted as $\mathbf{A}^{-1}$, is another n x n matrix B, so that $\mathbf{AB} = \mathbf{BA} = \mathbf{I}$, where I is the so-called identity matrix that contains 1's on its diagonal and 0's everywhere else. For example when $n = 2$, I is

$$\mathbf{I} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

The inverse of the general matrix A is calculated as follows:

$$\mathbf{A}^{-1} = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} = \frac{1}{\det(\mathbf{A})} \begin{pmatrix} A_{2,2} & -A_{1,2} \\ -A_{2,1} & A_{1,1} \end{pmatrix}$$

Where, det(A) is what is known as the *determinant* of A and is defined as

$$\det(\mathbf{A}) = \det\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} = \left|\begin{pmatrix} A_{1,1} & A_{2,1} \\ A_{1,2} & A_{2,2} \end{pmatrix}\right|$$
$$= A_{1,1}A_{2,2} - A_{1,2}A_{2,1}$$

To find the determinant of a matrix with dimensions larger than 2 x 2, we refer to textbooks on linear algebra [see, for example Strang (1988)]. It is important to note that the inverse of A does not exist when its determinant equals zero. Such matrices that cannot be inverted are commonly referred to as *singular matrices*. Note also that the transpose of $\mathbf{A}^{-1}$, *i.e.*, $(\mathbf{A}^{-1})^{T}$, is equal to $(\mathbf{A}^{T})^{-1}$.

Another important matrix property is the *rank* of a matrix (r), which corresponds to the *number of genuinely independent rows* in a matrix. For an n x n square matrix A, when $r = n$, it can be proven that (1) A has an inverse and (2) this inverse is unique.

Example 2

For the matrices given in Example 1, the determinants are calculated to be det(A) = -2 and det(B) = 2, which indicates that these are nonsingular matrices that should have an inverse. These are calculated as follows:

$$\mathbf{A}^{-1} = \begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix}^{-1} = \begin{pmatrix} -1.5 & 0.5 \\ 1 & 0 \end{pmatrix} \text{ and}$$

$$\mathbf{B}^{-1} = \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix}^{-1} = \begin{pmatrix} 1 & -1.5 \\ -1 & 2 \end{pmatrix}$$

Such 2 x 2 systems are rather easy to handle on paper; however, for matrices with larger dimensions, software packages such as MATLAB, MATHCAD, or MATHEMATICA should be employed to facilitate these operations.

where $E_c^{\#}$ is the so-called pseudo-inverse (or the Moore-Penrose inverse) of the matrix given by

$$E_c^{\#} = \left(E_c^T E_c\right)^{-1} E_c^T \tag{4.15}$$

Equation (4.14) is essentially the least squares estimate of the nonmeasured rates contained in the vector r_c, where all balances have been employed for their determination. It can be shown that if E_c has full rank (*i.e.*, there are at least as many linearly independent balances as the number of nonmeasured rates), then $E_c^T E_c$ also has full rank and the pseudo-inverse therefore can be found.

In the case of an overdetermined system, after the nonmeasured rates (r_c) have been determined by eq. (4.14), one may still be left with unused balances that can be employed to check the overall consistency of measured and calculated rates. To accomplish this, eq. (4.14) is inserted into eq. (4.11) to yield

$$\mathbf{R}r_m = 0 \tag{4.16}$$

with **R** being the so-called *redundancy matrix* (van der Heijden *et al.*, 1994a,b) given by

$$\mathbf{R} = E_m - E_c\left(E_c^T E_c\right)^{-1} E_c^T E_m \tag{4.17}$$

The rank of the redundancy matrix specifies the number of independent equations that must be satisfied by the measured and calculated [per eq. (4.14)] rates, and therefore it contains *I* - rank(**R**) *dependent* rows. If the dependent rows are removed, we obtain rank(**R**) independent equations relating the measured variables, *i.e.*

$$\mathbf{R}_r r_m = 0 \tag{4.18}$$

where $\mathbf{R}_r$ is the reduced redundancy matrix containing only the independent rows of **R**. Equation (4.16) is the basis for our further analysis of gross error identification, but before we proceed, we first illustrate the preceding concepts and the method of determining $\mathbf{R}_r$.

EXAMPLE 4.6

Analysis of Aerobic Yeast Cultivation without Ethanol Formation

We return to the case of aerobic yeast cultivation, which was also discussed in Examples 4.1-4.3, but now we consider the situation where

there is no ethanol formation. In this case Y_{xe} is zero, and therefore we will not include ethanol in the black box model. Thus, we have μ, r_c, r_w, r_s, r_o, and r_N as the rates in the black box model. With measurements of the specific glucose uptake rate, specific oxygen uptake rate, specific carbon dioxide production rate, and specific growth rate (equal to the dilution rate in a steady state chemostat), the elemental matrix is partitioned as follows:

$$\begin{matrix} & \text{Glc} & O_2 & CO_2 & \text{biomass} \end{matrix} \qquad \begin{matrix} NH_3 & H_2O \end{matrix}$$

$$\mathbf{E}_m = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 2 & 0 & 0 & 1.83 \\ 1 & 2 & 2 & 0.56 \\ 0 & 0 & 0 & 0.17 \end{pmatrix}; \mathbf{E}_c = \begin{pmatrix} 0 & 0 \\ 3 & 2 \\ 0 & 1 \\ 1 & 0 \end{pmatrix} \tag{1}$$

Note that the four columns of matrix $\mathbf{E}_m$ correspond to the four rates of glucose, oxygen, carbon dioxide, and biomass, respectively, whereas the two columns of the $\mathbf{E}_c$ matrix correspond to ammonia and water rates, respectively. The rows of these above matrices represent, of course, the four elemental balances. With a total of six compounds and four elemental balances there are $F = 2$ degrees of freedom. Because four rates are measured, the system is overdetermined. By using eq. (4.17) the redundancy matrix is found to be

$$\mathbf{R} = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & -0.286 & -0.286 & 0.014 \\ 0 & 0.572 & 0.572 & -0.028 \\ 0 & 0.858 & 0.858 & -0.042 \end{pmatrix} \tag{2}$$

with rank($\mathbf{R}$) = 2. It is easily seen that the last two rows of $\mathbf{R}$ are proportional to the second row (the third row is equal to the second row, multiplied by -2, and the fourth row is equal to the second row multiplied by -3). We therefore delete these two rows and thereby obtain the reduced redundancy matrix:

$$\mathbf{R}_r = \begin{pmatrix} 1 & 0 & 1 & 1 \\ 0 & -0.286 & -0.286 & 0.014 \end{pmatrix} \tag{3}$$

Equation 3, along with the four measured specific rates, yields the following redundant equations according to eq. (4.18):

$$\mathbf{R}_r \mathbf{r}_m = \begin{pmatrix} -r_s + r_c + \mu \\ 0.286 r_o - 0.286 r_c + 0.014\mu \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{4}$$

Obviously the first row is recognized as a carbon balance, but the second row is not that easily identified even though it contains all the information from the constraints of the three other elemental balances.

Normally, experimental data are overlaid with noise, and in some cases there may even be systematic errors. As a consequence of such errors, eq. (4.16) is not, in general, exact. There will be some residuals different from zero when the measured rates (or yields) are multiplied into the reduced redundancy matrix. This is better expressed by recognizing that the *measured* rate vector $\bar{\mathbf{r}}_m$ equals the sum of the *actual rate* vector $\mathbf{r}_m$ and its corrupting general measurement error δ:

$$\overline{\mathbf{r}_m} = \mathbf{r}_m + \delta \tag{4.19}$$

Combination eq. (4.19) into eq. (4.18) yields the following equation for the vector of the residual ε:

$$\varepsilon = \mathbf{R}_r\overline{\mathbf{r}_m} = \mathbf{R}_r(\mathbf{r}_m + \delta) = \mathbf{R}_r\delta \tag{4.20}$$

If the model is correct and if there are no systematic or random errors, *i.e.*, $\delta = 0$, all equations [eq. (4.18)] are satisfied exactly and yield zero values for the residuals. In all data sets, however, there is some noise present in the measurements that makes the residuals vector different from zero. The best rate estimates are those that minimize the magnitude of the residual, and they are determined as follows.

By assuming that the error vector is distributed normally with a mean value of zero and a variance-covariance matrix F,

$$E(\delta) = 0 \tag{4.21}$$

$$\mathbf{F} \equiv E\left[(\bar{\mathbf{r}}_m - \mathbf{r}_m)(\bar{\mathbf{r}}_m - \mathbf{r}_m)^T\right] = E(\delta\delta^T) \tag{4.22}$$

where E is the expected value operator, it can be shown that the residuals also will be distributed normally with a mean of zero

$$E(\varepsilon) = \mathbf{R}_r E(\delta) = 0 \tag{4.23}$$

and a variance-covariance matrix given by

$$\mathbf{P} = E(\varepsilon\varepsilon^T) = \mathbf{R}_r E(\delta\delta^T)\mathbf{R}_r^T = \mathbf{R}_r\mathbf{F}\mathbf{R}_r^T \tag{4.24}$$

The minimum variance estimate of the error vector δ is obtained by minimizing the sum of squared errors scaled according to their variance:

$$\min_{\delta} \left(\delta^T \mathbf{F}^{-1} \delta \right) \tag{4.25}$$

The solution is given by

$$\hat{\delta} = \mathbf{F}\mathbf{R}_r^T \mathbf{P}^{-1} \varepsilon = \mathbf{F}\mathbf{R}_r^T \mathbf{P}^{-1} \mathbf{R}_r \mathbf{r}_m \tag{4.26}$$

where the circumflex specifies that the value of δ is an *estimate*. Because δ is distributed normally, the function to be minimized in eq. (4.25) is the same for the least squares minimization problem and for the maximum likelihood minimization problem. If the error vector is not distributed normally, the estimate in eq. (4.26) remains valid for the least squares minimization problem but it no longer is the maximum likelihood estimate (Wang and Stephanopoulos, 1983). By using eq. (4.26), the best estimates for the measured rates are obtained as

$$\hat{\mathbf{r}}_m = \overline{\mathbf{r}_m} - \hat{\delta} = \left(\mathbf{I} - \mathbf{F}\mathbf{R}_r^T \mathbf{P}^{-1} \mathbf{R}_r \right) \overline{\mathbf{r}_m} \tag{4.27}$$

where I is a identity matrix. It can be shown that the estimates of the measured rates given by eq. (4.27) have a smaller standard deviation than the raw measurements (Wang and Stephanopoulos, 1983), and the estimate therefore is likely to be more reliable than the measured data. By using the best estimates for the measured rates, the nonmeasured rates of the black box model can be calculated using eq. (4.14).

EXAMPLE 4.7

Analysis of Aerobic Yeast Cultivation without Ethanol Formation (Continued)

We now continue our analysis of aerobic yeast cultivation that was initiated in Example 4.6, where we derived the reduced redundancy matrix. At a dilution rate of 0.15 h^{-1}, the measured specific rates for glucose, oxygen, carbon dioxide, and biomass are given by

$$\bar{\mathbf{r}}_m = \begin{pmatrix} -r_s \\ -r_o \\ \mu \\ r_c \end{pmatrix} = \begin{pmatrix} -0.250 \\ -0.113 \\ 0.113 \\ 0.141 \end{pmatrix} \tag{1}$$

with all rates in C-moles (C-mole biomass hour)$^{-1}$. We now want to calculate better estimates for the measured rates when it is assumed that there is a 5% error in the biomass and glucose measurements and a 10% error in the gas measurements, *i.e.*, the oxygen and carbon dioxide measurements. With these errors the variance-covariance matrix is given by

$$\mathbf{F} = 10^{-3}\begin{pmatrix} 0.1563 & 0 & 0 & 0 \\ 0 & 0.1277 & 0 & 0 \\ 0 & 0 & 0.0319 & 0 \\ 0 & 0 & 0 & 0.1988 \end{pmatrix} \tag{2}$$

and by using eq. (4.24) we find

$$\mathbf{P} = 10^{-3}\begin{pmatrix} 0.3870 & -0.0563 \\ -0.0563 & 0.0267 \end{pmatrix} \tag{3}$$

(the reduced redundancy matrix is taken from Example 4.6). The error vector for the measured fluxes is then found from eq. (4.26):

$$\hat{\delta} = \begin{pmatrix} -0.0055 \\ 0.0115 \\ -0.0013 \\ 0.0108 \end{pmatrix} \tag{4}$$

and this leads to better estimates for the measured fluxes:

$$\hat{\mathbf{r}}_m = \begin{pmatrix} -0.2445 \\ -0.1245 \\ 0.1143 \\ 0.1302 \end{pmatrix} \tag{5}$$

Thus, there are only small corrections to the measurements, and the original measurements therefore seem to be good. However, the corrected measurements conform better with the elemental balances, and therefore they are better estimates than the raw measurements.

Normally, the variance-covariance matrix is assumed to be diagonal, meaning that the measurements are uncorrelated. However, specific rates, yield coefficients, and even volumetric rates are seldom measured directly but instead are derived from measurements of the so-called prime variables, which may influence more than one measured rate. An example is the

measurement of the oxygen uptake rate and the carbon dioxide production rate, which are both based on the measurement of the gas flow rate through the bioreactor, in conjunction with measurements of the partial pressure of the two gasses in the head space. If there is an error in the measurement of the gas flow rate, this influences both rates, and, therefore, errors in the measured rates are indirectly correlated. The same objection holds for the measurement of other rates as well that are normally obtained by combination of concentration and flow rate measurements. In all such cases with indirect error correlations, it is difficult to specify the true variance-covariance matrix F. Madron *et al.* (1977) describe a simple algorithm by which the true variance-covariance matrix can be found when the properties of the noise of the prime variables are known (see Box 4.3). In many cases, however, we have inadequate information about the noise of even the prime variables, and the true variance-covariance matrix cannot be derived. In these cases, one may decide to neglect covariances and use a diagonal variance-covariance matrix, where reasonable values for the errors are used. Alternatively, one may use the least squares estimate given by

$$\hat{\mathbf{r}}_m = \left(\mathbf{I} - \mathbf{R}_r^T\left(\mathbf{R}_r\mathbf{R}_r^T\right)^{-1}\mathbf{R}_r\right)\bar{\mathbf{r}}_m \tag{4.28}$$

which is based on the assumption of the same absolute error in all the measured rates. Because absolute values for the errors are used, it is only reasonable to apply eq. (4.28) when the variables are of the same magnitude.

If any constraint residuals are *significantly different from zero*, either a systematic error is present in at least one of the measurements or the model employed is incorrect. To quantify the statement "significantly different from zero," we introduce the test function h given by the sum of *weighted* squares of the residuals:

$$h = \varepsilon^T \mathbf{P}^{-1}\varepsilon \tag{4.29}$$

When the raw measurements are uncorrelated, the test function h is χ^2 distributed (Wang and Stephanopoulos, 1983), and this was shown to be the case for correlated raw data as well (van der Heijden *et al.*, 1994b). The degrees of freedom of the χ^2 distribution are equal to the rank(**P**) = rank(**R**), *i.e.*, the number of independent constraints. By comparing the calculated value of the test function h with the values of the χ^2 distribution at the degrees of freedom [rank(**R**)], it is possible to detect the presence of a systematic error in the data at a certain confidence level. Thus, if at a high enough confidence level one obtains a test function value that is greater than the value of the χ^2 distribution, then there is something wrong with the data

BOX 4.3

Calculation of the Variance-Covariance Matrix from Errors in Prime Variables

Normally the measured rates are determined from measurements of the so-called prime variables, *e.g.*, the volumetric glucose uptake rate in a steady state chemostat is determined as the difference between the glucose concentration in the feed and that in the bioreactor multiplied by the dilution rate. Specification of the variance-covariance matrix therefore is not straightforward. Madron *et al.* (1977) describe a simple approach to find F from the measurement noise of the prime variables. First, the measured rates are specified as functions of the prime variables. When the latter are collected in the vector y, we have for the ith rate:

$$r_{m,i} = f_i(y) \tag{1}$$

Generally the functions f_i are nonlinear, but in order to obtain an approximate estimate of the variances and covariances, these functions are linearized. The error of the measured i'th rate, δ_j, is expressed as a linear combination of the errors δj* of the prime variables:

$$\delta_i = \sum_{j=1}^{k} \left(\frac{\partial f_i}{\partial y_j} \right) \delta_i^* = \sum_{j=1}^{k} g_{ij} \delta_j^* \tag{2}$$

where K is the number of prime variables and g_{ij} are the sensitivities. If the sensitivities are collected in the matrix G, the variance-covariance matrix F can be calculated from

$$\mathrm{F} = \mathrm{GF}^*\mathrm{G}^T \tag{3}$$

where F* is a diagonal matrix with the variances of the prime variables. The preceding method is very simple to compute the covariances, but obviously the calculated values are limited by the accuracy of the linear approximation in eq. (2).

TABLE 4.4 Values of the χ^2 Distribution

	Confidence level					
Degrees of freedom	0.500	0.750	0.900	0.950	0.975	0.990
1	0.46	1.32	2.71	3.84	5.02	6.63
2	1.39	2.77	4.61	5.99	7.38	9.21
3	2.37	4.11	6.25	7.81	9.35	11.30
4	3.36	5.39	7.78	9.49	11.10	13.30
5	4.35	6.63	9.24	11.10	12.80	15.10

or the model. Table 4.4 gives values of the χ^2 distribution at different confidence levels and different degrees of freedom.

EXAMPLE 4.8

Analysis of Aerobic Yeast Cultivation without Ethanol Formation (Continued)

We now continue our analysis of aerobic yeast cultivation that was initiated in Example 4.6 and analyzed further in Example 4.7. From the matrices derived in Example 4.7, we calculate the residuals using eq. (4.20):

$$\varepsilon = \begin{pmatrix} 0.0040 \\ -0.0064 \end{pmatrix} \tag{1}$$

and then the test function can be calculated using eq. (4.29):

$$h = \varepsilon^T \mathbf{P}^{-1} \varepsilon = 1.87 \tag{2}$$

Because there are two independent rows in the reduced redundancy matrix, its rank is 2, *i.e.*, the test function has to be compared with the χ^2 distribution with two degrees of freedom. From Table 4.4 it is seen that the test function is lower than the χ^2 distribution even at a confidence level of 0.75. Thus, it is only at a very low confidence level that it can be concluded that the data contain gross errors; hence, the data quality is satisfactory.

The finding of a large value for the test function, $h > \chi^2$, at a given confidence level does not allow one to conclude whether the unsatisfactorily large errors are due to systematic errors in the data or due to large random

errors. One approach that can be applied to this end is to eliminate one measurement at a time from the given set of data and use one of the constraints to calculate such a measurement. The remaining constraints are then used for consistency analysis to recalculate the test function h and compare it again with the χ^2 statistic for one fewer degree of freedom. If a significantly lower value is obtained for the test function upon elimination of a certain measurement, this is strong evidence for the presence of gross (systematic) errors in the measurement that was eliminated. The same can be applied to constraints other than the elemental balances, such as those arising from application of the steady state hypothesis to intracellular metabolites. This approach of error diagnosis requires that the system be overdetermined by at least two measurements, *i.e.*, rank(**R**) $\geq$ 2. This overdetermination allows for one constraint to be used for the calculation of the eliminated measurement while the other is used for the recalculation of the test function. The procedure of measurement elimination is very simple, as illustrated in Example 4.9, and it allows for the rapid determination of the probable source of a systematic error.

EXAMPLE 4.9

Error Diagnosis in Yeast Cultivation Measurements

For aerobic growth of *S. cerevisiae* with glucose as the carbon source and ammonia as the nitrogen source, the specific rates of glucose uptake, oxygen uptake, biomass growth, and carbon dioxide formation are measured to be

$$\mathbf{r} = \begin{pmatrix} -r_s \\ -r_o \\ \mu \\ r_c \end{pmatrix} = 0.008 \begin{pmatrix} -2.1 \\ -3.8 \\ 1 \\ 1.4 \end{pmatrix} \tag{1}$$

at a specific growth rate of 0.008 h^{-1} [all rates in C-moles (C-mole biomass hour)$^{-1}$]. The elemental composition of the biomass is assumed to be the same as in Examples 4.6-4.8. The measurement errors are 6%, 11.7%, 5%, and 11.1% for glucose, oxygen, biomass, and carbon dioxide, respectively. There are no covariances. We now want to examine whether there are any experimental errors. Because the stoichiometry is the same and the same rates are measured as in Examples 4.6-4.8, we can use the reduced redundancy matrix derived in Example 4.6. Furthermore, with the given errors the

variance-covariance matrix is found to be

$$\mathbf{F} = 10^{-4}\begin{pmatrix} 0.0102 & 0 & 0 & 0 \\ 0 & 0.1265 & 0 & 0 \\ 0 & 0 & 0.0016 & 0 \\ 0 & 0 & 0 & 0.0155 \end{pmatrix} \tag{2}$$

When the test function is calculated as illustrated in Example 4.8, it is found to be

$$h = 35.06 \tag{3}$$

which shows that even at a confidence level of 0.99 there is a measurement error. From inspection of the measured rates it seems likely that this error is in either the oxygen or the carbon dioxide measurement, as the respiratory quotient ($RQ = r_c/r_o$) is less than the 1, the normal value for an aerobic culture of *S. cerevisiae* growing at low specific growth rates. To identify the measurement error, we do, however, eliminate each of the four reactions - one at a time - and then calculate the test function. The result of this gives

Compound eliminated	*h*
Glucose	27.06
Oxygen	2.12
Biomass	26.43
Carbon dioxide	34.96

Clearly, if any of the three measurements glucose, biomass, or carbon dioxide is eliminated, there is still a measurement error. Only when the oxygen measurement is eliminated does the value of the test function drop to a low value, and by comparing it with the χ^2 distribution with one degree of freedom it is seen that at a confidence greater than 90% it cannot be concluded that there are gross measurement errors. Thus, it is very likely that the oxygen measurement is erroneous.

If the oxygen measurement is left out, it is possible to calculate both better estimates for the three measured rates and best estimates for the three nonmeasured rates (including oxygen). First, by using eq. (4.27), we find for the measured rates

$$\hat{\mathbf{r}}_m = \begin{pmatrix} -r_s \\ \mu \\ r_c \end{pmatrix} = 0.008\begin{pmatrix} -2.21 \\ 0.98 \\ 1.23 \end{pmatrix} \tag{4}$$

and thereafter we find the nonmeasured rates (ammonia uptake, water formation, and oxygen uptake) using eq. (4.14) [with eq. (4) inserted for the

measured rates]:

$$\hat{\mathbf{r}}_c = \begin{pmatrix} -r_N \\ r_w \\ -r_o \end{pmatrix} = 0.008 \begin{pmatrix} -0.17 \\ 1.56 \\ -1.18 \end{pmatrix} \tag{5}$$

Thus, we see that the oxygen uptake rate is corrected drastically. Furthermore, with the estimated rates the RQ is found to be 1.04, which is a much more realistic value.

REFERENCES

Abbott, B. J. & Clamen, A. (1973). A. The relationship of substrate, growth rate, and maintenance coefficients to single cell protein production. *Biotechnology and Bioengineering* 15, 117-127.

Christensen, L. H. & Henriksen, C. M.; Nielsen, J. & Villadsen, J. (1995). Continuous cultivation of *Penicillium chrysogenum*. Growth on glucose and penicillin production. *Journal of Biotechnology* 42, 95-107.

Erickson, L. E., Minkevich, I. G. & Eroshin, V. K. (1978). Application of mass and energy balance regularities in fermentation. *Biotechnology and Bioengineering* 20:1595-1621

Larsson, C., Blomberg, A. & Gustafsson, L. (1991). Use of microcalorimetric monitoring in establishing continuous energy balances and in continuous determinations of substrate and product concentrations of batch grown *Saccharomyces cerevisiae*. *Biotechnology and Bioengineering*. 38, 447-458.

Madron, F., Veverka, V. & Vanecek, V. (1977). Statistical analysis of material balance of a chemical reactor. *AIChE Journal* 23, 482-486.

Nielsen, J. & Villadsen, J. (1994). *Bioreaction engineering principles*. New York: Plenum Press.

Roels, J. A. (1983). *Energetics and Kinetics in Biotechnology*. Amsterdam: Elsevier Biomedical Press.

Schulze, U. (1995). Anaerobic physiology of *Saccharomyces cerevisiae*. Ph.D. Thesis, Technical University of Denmark.

van der Heijden, R. T. J. M., Heijnen, J. J., Hellinga, C., Romein, B. & Luyben, K. Ch. A. M. (1994a). Linear constraint relations in biochemical reaction systems: I. Classification of the calculability and the balanceability of conversion rates. *Biotechnology and Bioengineering* 43, 3-10.

van der Heijden, R. T. J. M., Heijnen, J. J., Hellinga, C., Romein, B. & Luyben, K. Ch. A. M. (1994b). Linear constraint relations in biochemical reaction systems: II. Diagnosis and estimation of gross measurement errors. *Biotechnology and Bioengineering*. 43, 11-20.

von Meyenburg, K. (1969). Katabolit-Repression und der Sprossungszyklus von *Saccharomyces cerevisiae*. Ph.D. Thesis, ETH Zürich.

von Stockar, U. & Birou, B. (1989). The heat generated by yeast cultures with a mixed metabolism in the transition between respiration and fermentation. *Biotechnology and Bioengineering*. 34, 86-101.

Wang, N. S. & Stephanopoulos, G. (1983). Application of macroscopic balances to the identification of gross measurement errors. *Biotechnology and Bioengineering*. 25, 2177-2208.