DELFT UNIVERSITY OF TECHNOLOGY

ANALYSIS OF METABOLIC NETWORKS
COURSECODE

Module 1: Rates, Reconciliation, Black-box modeling

Authors:Sef Heijnen, Aljoscha Wahl, Timmy Paez Watson
November 9, 2023

Contents

1 Rates and balances

1.1 Importance of rates

Living systems transform matter and energy with certain rates (mol_i/h and kJ/h). Knowing these rate values and how these can be changed is important.

- In *industrial processes,* these rates determine equipment **design** (pumps, stirring power etc.) and **economy** (profits, investments).
- In *natural environments* these rates determine e.g. how fast pollutants are transported, degraded and accumulated in higher organisms (animals) (leading to **adverse effects**).
- In *our bodies*, **health problems** from metabolic disorders quickly result from imbalances in rates of production, transport, and degradation of metabolites or proteins causing too high or too low levels in certain parts of our bodies, usually with fatal results.

In all these applications cells and **(micro)organisms consume** the nutrients made available in the growth medium and **produce** other compounds at certain rates. **Quantification of these rates** is of high importance.

Moreover, it is important to understand how a cell / microorganism **changes its rates in response to a change in its extra-cellular environment**, which is defined as temperature, pH, concentration of substrate, product and of many other compounds outside the cell membrane.

Rates and health

Drugs are degraded (in e.g. liver) and secreted (kidney). Information on the rate of conversion of a drug in the liver and of the rate of secretion in the kidney is needed to calculate the drug supply needed to maintain a desired drug level in e.g. the blood of a patient. This area is called pharmacokinetics. Some drugs are administered over long times e.g. birth control hormones and nicotine patches against cigarette addiction. The design of these systems is based on knowing the rate of transport through the skin and the rate of degradation of the drug.

1.2 Conversion rate R_i of compound i must be calculated from its balance

It is known that during a biological process, many different compounds are involved: consumed electron donor and electron acceptor, produced biomass and product, consumed or produced protons, consumed N-source, produced or consumed CO_2, water etc.

In an experiment (or in a factory, in your body, etc.) each of these compounds is consumed or produced with a certain rate. These **conversion rates** are usually expressed as **amount per time**, (e.g. **mol of compound i per hour)** and given the symbol $R_i mol(i)/h$.

One might think that one can measure these rates using a certain sensor, similar to the use of a thermometer to measure temperature, or a pH-electrode to measure the pH etc. Fundamentally, this is impossible. Hence, we **cannot measure conversion rates** with a sensor, such a device **to measure a rate of production or consumption does not exist**.

The conversion rate of a compound i (mol_i/h) **must be calculated** using the **balance for compound i** in combination with **proper measurements** of flow rates, volumes, concentrations and time (in our body, in a fermentation process, in nature or in our experiments).

1.3 Definition of the balance

For **each compound** i one must formulate a **balance** (formally called "**compound balance**).

Each balance has the **same general structure based on the following statement**:

“the amount (mol_i) of **compound** i, present within a **defined system boundary**, changes in time (in mol_i/h) as a result of **transport** of compound i into or out of that defined volume **and** of **transformation** (production and/or consumption) of compound i because of (bio)chemical reactions present in that system."

This results in the **balance equation of compound i**

$$\frac{d(V_L c_i)}{dt} = R_i + F_{in}c_{i,in} - F_{out}c_{i,out}$$

Accumulation/ depletion of compound i	Rate of production/ consumption of compound i	Transport of compound i
[mol i/h]	[mol i/h]	[mol i/h]

It is important to note that this is a **balance of rates related to compound** i and that each term in this balance has the same dimension: mol_i/h.

It is also obvious that the balance of compound i allows us to **calculate** R_i **provided that** we can **quantify all the other terms in the balance using experimental measurements**.

1.3.1 A balance always needs a system boundary

A balance always needs a defined system boundary. **Such boundaries are chosen in a case-specific practical way.** For example, if one is interested in uptake rates by all cells present in the fermentor **broth** ($=$ aqueous suspension containing the organisms) the logical system boundary is the envelope of the whole broth present in the fermentor. If one is interested in the transfer of O_2 from the gas phase from lungs to blood, the system boundary comprises the whole gas volume in the lungs. Obvious reasons to **choose** boundaries are e.g. different phases (gas/liquid/solid) or compartments in equipment or in living systems.

1.3.2 Computing balances requires knowledge about the mechanisms

When a (bio)chemical system is studied one usually observes that concentrations change in time. **Understanding why the concentration of compound** i **changes** requires a thorough understanding of the **underlying mechanisms which work on compound** i**.**

Note that when concentration c_i **does not change in time, this does not mean that mechanisms are absent.**

The key is to ask oneself “Which **mechanisms** act **simultaneously** on the compound i. This includes **two types of mechanisms: production/consumption** and **transport**". The combined result of **the positive and negative rates of these mechanisms** determines the increase, decrease **or constancy of the amount** ($< 0, < 0, = 0$) of i (mol_i/h) present within the boundary of the system, as described in the balance of compound i.

1.3.3 System schemes and notation used in balances

A practical way to define balances is to make a sketch (system scheme) and use certain symbols.

The first step is to draw a scheme which represents the considered system and its boundaries.

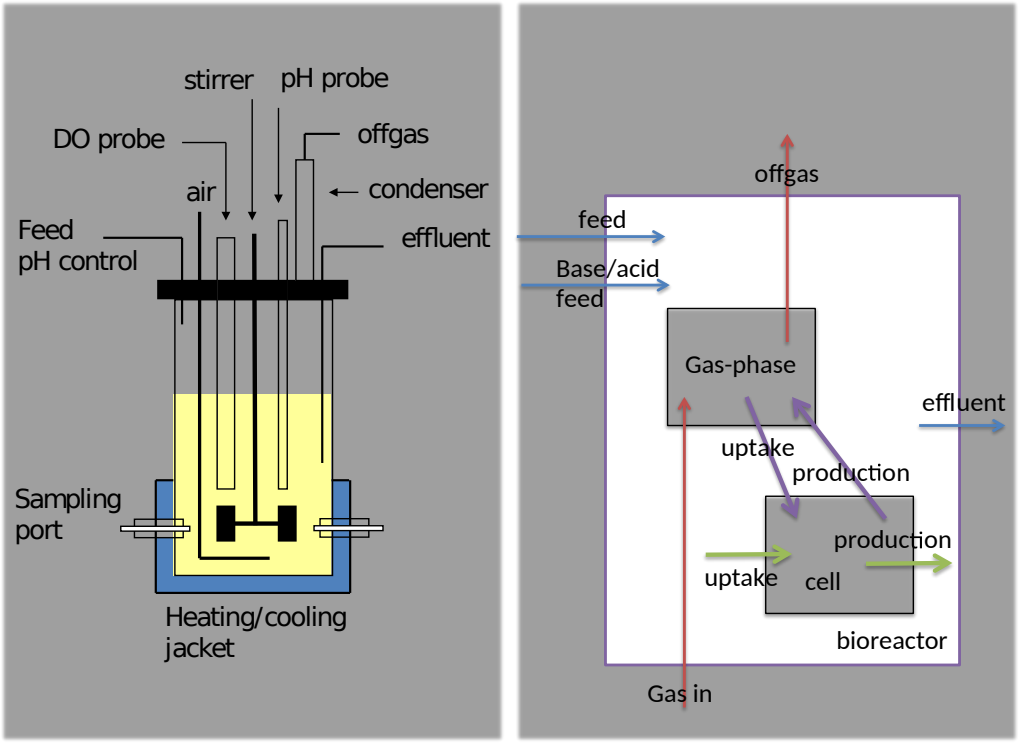

Figure 1: System schemes and transport between balance spaces.

Shown to the left is the system scheme in the case that the compound of interest is only present in the broth in a fermentor, e.g. biomass or **non-volatile product**. The system then contains all the broth present in the fermentor.

To the right we have **two connected systems,** the gas phase and the broth phase which are present in the fermentor, each with its own amount (broth volume V_L for broth phase, total amount of gas molecules N_G for the gas phase). A gas phase is required in addition to the broth phase when the molecule of interest is present in these two phases (e.g. O_2 or CO_2 or a **volatile substrate or product**).

The next step is to characterize the **amount (mol_i) of compound** present in (each of) the defined system(s).

When the system represents a **liquid,** we can express the amount (mol_i) of compound in liquid as:

$$N_i = V_L c_i \tag{1}$$

where c_i is the concentration of compound i present in the liquid phase and V_L is the volume of said liquid phase.

When the system represents a **gas** we express the amount (mol_i) of compound i in gas as:

$$N_i = N_G y_i \tag{2}$$

where N_G is the total amount of molecules present in the gas phase within the system, y_i is the mole fraction of the compound i in the gas phase.

NOTE: About N_G and y_i in the gas phase

N_G is the total amount (mol) of gas molecules present in the defined gas phase, e.g. the gas present in all gas bubbles in your fermentor. N_G cannot be measured as such, but can be calculated using the ideal gas law and the measured gas volume V_G, temperature, and pressure.

$$N_G = \frac{PV_G}{RT} \tag{3}$$

P is the measured pressure (N/m^2) in the gas

V_G is the measured volume of all gas bubbles (m^3)

R is the gas constant $(8.314 J/molK)$

T is the absolute temperature (K) in the gas phase in the fermentor

With respect to the mole fraction y_i this is usually directly obtained from measurements, where, according to the ideal gas law, volume fractions are identical to mole fractions.

1.4 Mechanisms and their rates

We can generally distinguish **two types of mechanisms**.

1.4.1 Type 1: Production (positive rate) or consumption (negative rate) mechanisms

In (biological) processes, cells and organism consume nutrients (e.g. carbon or nittrogen source, O_2 etc.) and produce new cells, (by)products, CO_2, heat, etc. For **each compound we define a rate** R_i in mol_i/h, where i represents the compound (s for substrate, o for O_2, x for produced cells, c for CO_2, n for nitrogen source, w for water, q for heat, h for H^+).

R_i can be positive (when compound i is produced) or negative when consumed. In biological processes this occurs usually in the liquid (broth) phase as indicated in the flow schemes (because cells need an aqueous environment).

1.4.2 Type 2: Transport mechanisms

In many situations compounds are transported **into**(+) or **from** $(-)$ the system using mechanisms which bring the compounds **over the system boundary** (Fig. 2).

There are two **transport mechanisms** which frequently occur (but there are more mechanisms possible).

- Convective transport This mechanism usually involves **flows of liquid** F (in m^3/h), which enter (F_{in}) or leave (F_{out}) the **liquid phase** (present within the system boundary). In the liquid flow a dissolved compound i is present at concentration $c_{i,in}$ or $c_{i,out}$. (Fig. 2, bottom panel)

  The **transport rates** (mol_i/h) by the **convective liquid flow transport mechanism** are then

  $F_{in}c_{i,in}$ when **in**, or $F_{out}c_{i,out}$ when **out**

  Alternatively, there can be convective transport using a **flow of gas** F_n (in mol total gas/h). This gas flow then **enters** ($F_{n,in}$) or leaves ($F_{n,out}$) the **gas phase** present in the system (Fig. 2, upper panel). In the gas stream compound i (e.g. O_2, CO_2, volatile product, evaporated H_2O) is present at **mole fraction** y_i. The transport **rates** (mol_i/h) by **convective gas flow** are then

  $F_{n,in}y_{i,in}$ when **in**, or $F_{n,out}y_{i,out}$ when **out**

- Transport by transfer

The above discussed convective transport occurs **within one phase** (liquid or gas). In contrast transfer occurs **between two phases**. Well known examples are transfer of O_2 **from gas bubbles to liquid** and CO_2-transfer in the **reverse direction**. The symbol for molar rate of transfer of compound i is $T_{n,i}$ (mol$_i$/h).

One should note that F indicates a liquid volume flow rate (m^3/h) whereas F_n and T_n represent molar flow rates (mol/h) The subscript n always indicates that the unit is mol.

The schemes for broth and broth/gas systems with the **complete notation for amounts present and rates of mechanisms (conversion, convective transport and transfer)** is shown below (Fig. 2).

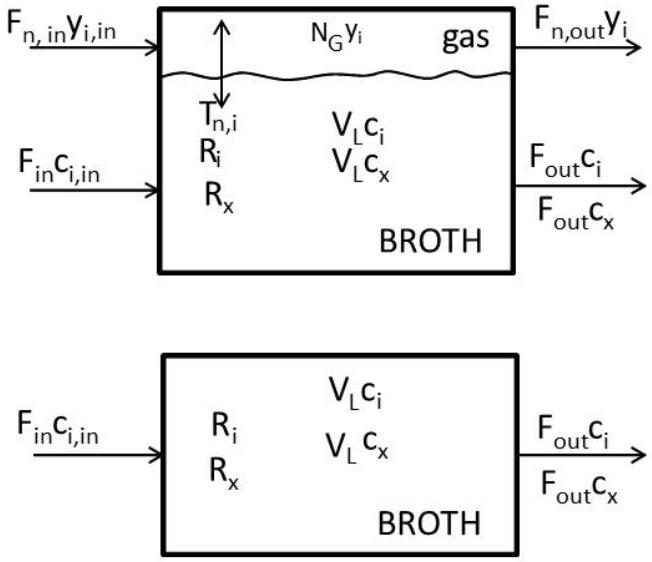

Figure 2: Symbols, rates of transport and conversion (all in mol, L or moli/h)

1.5 Balances

The following balances are now obtained from Fig. 2 based on the proposed system schemes and symbols:

Balance of compound i in broth (moli/h)

$$\frac{\mathrm{d}(V_L(t)c_i(t))}{\mathrm{d}t} = R_i(t) + F_{in}c_{i,in} + F_{out}(t)c_{i,out}(t) \pm T_{n,i}(t) \tag{4}$$

Balance of compound i in gas (moli/h)

$$\frac{\mathrm{d}(N_G(t)y_i(t))}{\mathrm{d}t} = 0 + F_{n,in}y_{i,in} + F_{n,out}(t)y_{i,out}(t) \pm T_{n,i}(t) \tag{5}$$

We see that there is no conversion (first term in zero) in the gas phase (**cells do not fly!!**).

NOTE: Number of balances We see from Fig. 2 and the above balances that when a compound i is present in 2 subsystems (e.g. broth and gas) there are 2 balances for that compound. This can be generalized, the number of balances being always equal to:

$$\text{Number of balances} = \begin{bmatrix}\text{Number of}\\ \text{compounds}\end{bmatrix} \times \begin{bmatrix}\text{Number of}\\ \text{subsystems}\end{bmatrix} \tag{6}$$

1.5.1 Non ideal outflow: What it is and how to avoid it

In the schemes of Fig. 2 we have distinguished **concentrations in outflows** $c_{i,out}$ and $y_{i,out}$ **which can be different** from the concentrations in the liquid and gas phase (c_i, y_i) present in the system. There are several reasons for this.

Liquid phase. When a compound is completely dissolved then $c_{i,out} = c_i$. When not (e.g. biomass, crystallized product, a non-soluble liquid product) then **special precautions/constructions regarding the outflow system** are needed to ensure that $c_{i,out} = c_i$. When such precautions are not taken then $c_{i,out} \neq c_i$.

For example, in case of liquid outflow via an overflow device, biomass concentrations in the outflow can be higher (hydrophobic biomass) or lower (hydrophilic biomass) than the biomass concentration in the fermentor. Similar differences in concentrations can occur for non-soluble solid or liquid products or substrates. Clearly, although a liquid overflow is a cheap and simple outflow device, it is very risky. A **clear recommendation** is to always measure both concentrations (c_i and $c_{i,out}$) and see whether there is a difference. When there is a difference there are many technical solutions for an outflow construction to avoid this situation. Good ideas are to

- Construct the outflow opening submerged in a well-mixed environment and use a pump to create outflow of well mixed broth.
- Install a suction pipe which dips at the broth surface from the top and use a large pump which operates intermittent and sucks very fast the broth upwards into an outflow container.

When this non ideal outflow situation cannot be corrected one needs to measure both c_i and $c_{i,out}$.

Gas phase. When the outflow gas is cooled to condense e.g. H_2O or a volatile product, then this leads to $y_{i,out} \neq y_i$. Usually $y_i = y_{i,out}$.

1.5.2 Experimental design: Balances reveal what you need to measure before the experiment

In many situations one needs to obtain the values for R_i for compound i in time from experiments. **It is strongly recommended to set up balances to identify the needed measurements BEFORE one does the experiments**.

The above **broth balance for compound** i shows that, to obtain the **conversion rate** R_i one needs to measure as function of time F_{in}, $c_{i,in}$, F_out, $c_{i,out}$, V_L, c_i **and the transfer rate** $T_{n,i}$ **(mol$_i$/h)**. This **transfer rate cannot be measured** by a sensor and it follows that **one must obtain the transfer rate** $T_{n,i}$ from a balance as for any rate. This balance turns out to be the **gas phase balance of compound i in the gas phase.** This gas balance shows that one, in addition, needs to measure as function of time $F_{n,in}$, $y_{i,in}$, $F_{n,out}$, $y_{i,out}$, N_G and y_i.

1.5.3 Ideal mixing in gas and liquid phase and ideal liquid outflow

In practice one often aims for **ideal mixing** (both gas and liquid phase) and **ideal broth outflow**. In the scheme below Fig.3 ideal mixing is indicated by the **propeller**.

In such an ideal situation, the concentrations in the broth and the broth outflow and gas phase/gas outflow are everywhere the same, hence $c_{i,out} = c_i$. Also for the gas phase then $y_{i,out} = y_i$.

The system scheme, notation and balances now become:

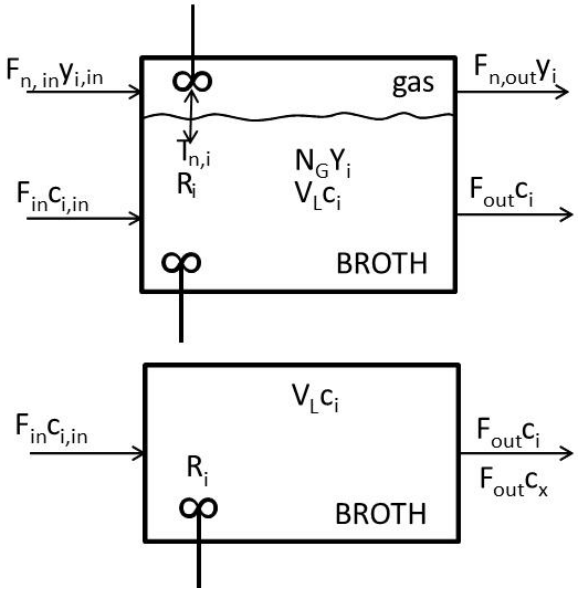

Figure 3: Symbols, conversion and transport rates (mol_i/h) for ideal (mixing and broth outflow) situation

The balances are then simplified (note that $c_{i,out}$ is now only c_i):

Balance of compound i in broth (mol_i/h)

$$\frac{\mathrm{d}(V_L(t)c_i(t))}{\mathrm{d}t} = R_i(t) + F_{in}c_{i,in} + F_{out}(t)c_i(t) \pm T_{n,i}(t) \tag{7}$$

Balance of compound i in gas (mol_i/h)

$$\frac{\mathrm{d}(N_G(t)y_i(t))}{\mathrm{d}t} = 0 + F_{n,in}y_{i,in} + F_{n,out}(t)y_i(t) \pm T_{n,i}(t) \tag{8}$$

The ideal situation diminishes the number of measurements needed to obtain R_i. But we still need the gas phase balance to obtain the **transfer rate** $T_{n,i}$ in case of volatile compounds.

1.5.4 Setting up the balances: some guidelines

For a process involving many chemical compounds and many phases (liquid, gas, solid) or compartments, one needs to formulate **many balances.**

- **There is one balance for each compound i in each phase or compartment.**
- **Each compound has its own concentration in each compartment or phase, and each compartment or phase has its own volume (or total mol present when gas phase).**
- Furthermore, one needs accurate information on the **transport mechanisms** (convective, transfer) which operate on each compound i in each phase in order to quantify the transport terms in each balance.
- Finally one has to specify the **conversion mechanisms** which operate on each compound and their rates.

1.6 $R_i(t)$ from balances in a batch

Having introduced the concept of balances allows us to address the question of obtaining the **rates**R_i(t) in the context of experiments.

A usual task is to cultivate cells/organisms and to study the rate of growth or product formation or some other rate. The experimenter is usually free to choose the method of cultivation. This choice is called “experimental design". The simplest method is called “**Batch cultivation**", which is executed in shake flasks, batch fermenters or other devices.

The characteristic of a *true* batch experiment is that, for *all* compounds involved in the biological process, there are **no mechanisms to transport these compounds into or from the broth present in the cultivation vessel: transport is absent (see note)**.

Traditionally, when one speaks of batch experiments this is usually assumed to imply that the volume is constant. In practice, a constant volume can seldom be realized. Firstly, reactions lead in principle to changes in densities. Also, losses often occur such as evaporation of water, or additions are applied (e.g. for pH control). Therefore, **a constant volume does not characterize a batch condition**. The key is **the absence or presence of transport, which must be specified for each molecule**.

For example, in an aerated (air sparging) batch fermentation, the batch condition applies for biomass and substrate present in broth, but not for O_2 and CO_2 in the broth, which are transferred between air bubbles and broth. When the pH is controlled with NH_4OH solution, the batch condition also does not apply to NH_4^+ in the broth. If the pH is controlled with NaOH, then the batch condition does apply for NH_4^+, but not for Na^+! This shows that in a batch experiment only for some compounds the balances relate to batch conditions, because there is no transport of these compounds. For other compounds the batch condition does not apply, because there are transport mechanisms present.

Now, considering those components that *are* in batch condition, the **balance for a batch compound i** is simplified (because the transport terms in and out are **absent**) to

$$\frac{\mathrm{d}(V_L(t)c_i(t))}{\mathrm{d}t} = R_i(t) \tag{9}$$

It is also useful to introduce $N_i = V_L c_i$, with N_i being the total amount (moles) of compound i present in the broth. So:

$$\frac{\mathrm{d}(V_L(t)c_i(t))}{\mathrm{d}t} = \frac{\mathrm{d}N_i(t)}{\mathrm{d}t} = R_i(t) \tag{10}$$

Note that R_i will be **positive** for a produced compound, and **negative** for a consumed compound.

From this balance, it is now very clear that **to obtain** $R_i(t)$ **in a batch experiment we have to measure** V_L **and** c_i **as function of time**, which will allow us to obtain $N_i(t) = V_L(t)c_i(t)$, which is the amount (mol) of i in the experiment, changing as function of time. The obtained experimental $N_i(t)$ values can be plotted as function of time, and the slope can be calculated. **In batch, this slope equals** R_i**(t)** according to the above balance.

The simplest way to calculate the slope of $N_i(t)$ at time t is using two two values of N_i at two points close in time:

$$\frac{\mathrm{d}N_i(t)}{\mathrm{d}t} = \frac{N_i(t_2) - N_i(t_1)}{t_2 - t_1} \tag{11}$$

Note that **in batch experiments the slope is typically not constant, but changes in time**. Thus, also more complicated methods of computing that changing slope, e.g. regression or curve fitting, can be applied.

The example below shows the use of **balances and experimental measurements** to calculate the values of R_i as a function of time from a batch experiment where an organism does grow and consumes glucose.

Work with the example

You can work with the following example in the Python Jupyter notebook file `L01_Batch.ipynb` that you can download on Brightspace. You will also need to download the excel file with the data, at `Data_Batch.xlsx`.

This example is also solved in the Lecture video (AMN module 1 part 1) available on Brightspace.

Example 1.1: R_i calculations from balances and measurements in a batch fermentation

An organism is grown on glucose in a batch fermenter. We want to calculate $R_x(t)$ and $R_s(t)$. $R_x(t)$ is obtained from the biomass balance and $R_s(t)$ is obtained from the substrate balance. Setting up these balances shows **what we need to measure:**

$$\frac{\mathrm{d}(V_L(t)c_x(t))}{\mathrm{d}t} = R_x(t) \qquad Cmol_x/h \tag{12}$$

$$\frac{\mathrm{d}(V_L(t)c_S(t))}{\mathrm{d}t} = R_x(t) \qquad mol_S/h \tag{13}$$

These balances show that, to obtain these rates as function of time, we need to measure $V_L(t)$, $c_x(t)$, $c_s(t)$.

The experiment is started and the measurement results shown below are obtained over a period of 16 hours. We see that the volume decreases, which is due to the evaporation of water because of air sparging. The evaporation rate is 2 litres of water per hour.

Time (h)	V_L (m^3)	c_S (mol/m^3)	c_x ($Cmol/m^3$)
0	0.100	100.0	100
1	0.098	100.3	107
2	0.096	100.5	115
4	0.092	100.7	133
8	0.084	99.5	178
16	0.068	87.0	327

Question 1: Explain why the **biomass concentration increases and the substrate concentration increases** between 0 and 4 hours. Remember that one expects glucose consumption for growth.

Answer 1: This is due to the volume decrease from 0.100 m^3 to 0.092 m^3.

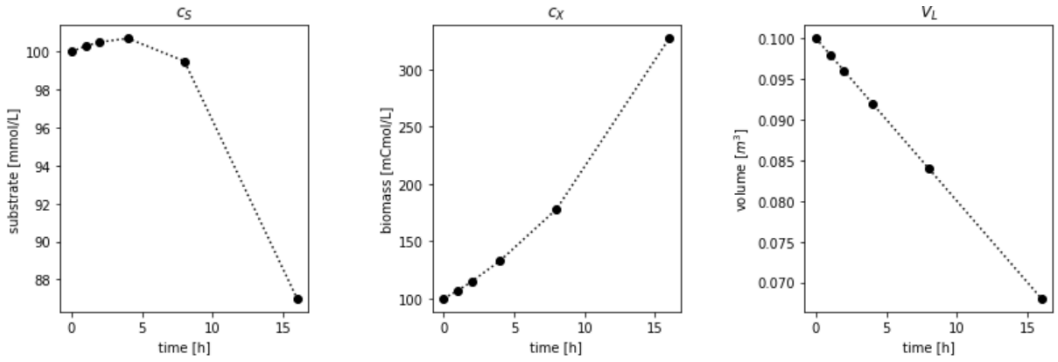

Figure 4: Measured concentration substrate, biomass and volume as function of time in a batch fermentation

Question 2: Calculate the rate of substrate consumption ***$R_s(t)$*** and biomass production ***$R_x(t)$*** (in mol/h and Cmol/h) as function of time (rather than using concentration).

Answer 2: We need to use the balance for glucose to obtain R_s and the balance for biomass to obtain R_x (as defined above).

We need **first to calculate the total amounts of glucose ($V_L c_s = N_s$) and biomass ($V_L c_x = N_x$),** (from the available volume and concentration measurements) and see how these total amounts change with time.

Time (h)	$V_L c_s = N_s$ (mol)	$V_L c_x = N_x$ (Cmol)
0	10.00	10.0
1	9.82	10.51
2	9.64	11.05
4	9.26	12.21
8	8.36	14.92
16	5.91	22.26

According to the batch balance, R_i is the slope of the curve of total amount of i, ($V_L c_i$) versus time. Ideally, one would need to obtain a continuous interpolation function through the ($V_L c_i$) -time points and then obtain the slope as function of time. For simplicity, we can approximate the slope by using the ($V_L c_i$) data points and the **discrete** definition of the slope in **the time interval t_1 to t_2**.

$$\frac{\mathrm{d}N_i(t)}{\mathrm{d}t} = \frac{(V_L(t_2)c_i(t_2)) - (V_L(t_1)c_i(t_1))}{t_2 - t_1} \tag{14}$$

For the available time intervals, we thus obtain the following values for the slopes at different time intervals (which are equal to the **average** R_i in the respective time interval):

Time interval (t1 to t2)	$d(V_L c_s)/dt$ (mol/h)	$d(V_L c_x)/dt$ (Cmol/h)
0-1 h	-0.17	+0.49
1-2 h	-0.18	+0.55
2-4 h	-0.19	+0.60
4-8 h	-0.23	+0.68
8-16 h	-0.31	+0.91

Now it is clear that there is growth (N_x increases) at the expense of glucose consumption (N_s decreases).

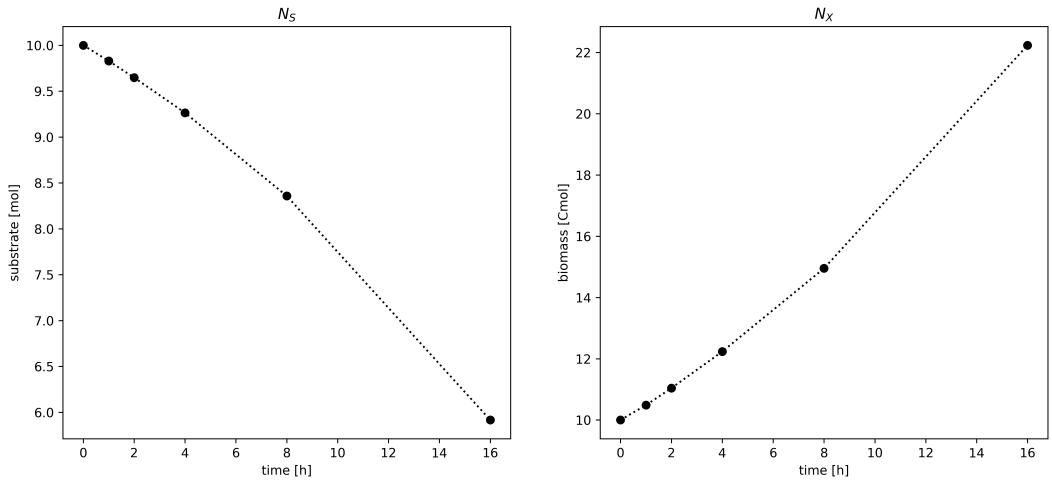

Figure 5: Amount of substrate ($V_L c_s$) and biomass ($V_L c_x$) as function of time in a batch fermentation

Python codes for these figures and tables (also available on Brightspace: `L01a-Batch`):

```
# import relevant libraries
import numpy as np
import numpy.linalg as linalg
import pandas as pd
import os
import matplotlib.pyplot as plt

# Get the data from the excel sheet

cwd = os.getcwd()
path = cwd + '\\Data_Batch.xlsx'
df = pd.read_excel(path, 'data', header=[0,1] )
data_meas = np.array(df)

# Assign data from excel sheet to variables
t   = data_meas[:,0]
V_L = data_meas[:,1]
c_S = data_meas[:,2]
c_X = data_meas[:,3]

# plot raw data
plt.figure( figsize = (12, 4 ))
plt.subplot(1, 3, 1)
plt.plot( t, c_S, 'ko:' )        # Substrate concentration change in time
plt.xlabel( 'time [h]' )
plt.ylabel( 'substrate [mmol/L]')
plt.title( '$c_S$' )

plt.subplot(1, 3, 2)
plt.plot( t, c_X, 'ko:' )        # Biomass concentration change in time
plt.xlabel( 'time [h]' )
plt.ylabel( 'biomass [mCmol/L]')
plt.title( '$c_X$' )

plt.subplot(1, 3, 3)
plt.plot( t, V_L, 'ko:' )        # Volume change in time
plt.xlabel( 'time [h]' )
plt.ylabel( 'volume [$m^3$]')
plt.title( '$V_L$' )
```

```
plt.subplots_adjust( right= 1., wspace = 0.5)
#plt.savefig( 'plt_exp_data.png', dpi=600)
plt.show()
```

```
# now calculate amounts V_L * c_i
N_X = V_L * c_X
N_S = V_L * c_S

# plot data
plt.figure( figsize = (8, 4 ))
plt.subplot(1, 2, 1)
plt.plot( t, N_S, 'ko:' )
plt.xlabel( 'time [h]' )
plt.ylabel( 'substrate [mol]')
plt.title( '$N_S$' )

plt.subplot(1, 2, 2)
plt.plot( t, N_X, 'ko:' )
plt.xlabel( 'time [h]' )
plt.ylabel( 'biomass [Cmol]')
plt.title( '$N_X$' )

#plt.savefig( 'plt_exp_data_N.png', dpi=600)
plt.show()
```

```
# calculate time interval differences (d(VL*cX)/dt)
dXdt = np.diff( N_X ) / np.diff( t )

# calculate time interval differences (d(VL*cS)/dt)
dSdt = np.diff( N_S ) / np.diff( t )

# print table with results dX/dt dS/dt

print( "time  dSdt  dXdt")
print( "-"*18)
for i in range( np.size( dXdt) ):
print( "{:2.0f}-{:2.0f} {:5.2f} {:5.2f}".format( t[i], t[i+1], dSdt[i], dXdt[i]) )
```

2 q-Rates

2.1 Definition

In the example in the previous section we have seen how in a batch cultivation R_s and R_x can be obtained from the balances for substrate and biomass using volume and concentration measurements as function of time.

The obtained result merits now some thought. The example shows (see Figs.4 and 5 in Example 1) that the rate of glucose consumption **(-R_s) increases in time**. **This seems strange**; at first glance, one would expect that the consumption rate of glucose would decrease when the glucose amount decreases (left panel Fig.4).

The **explanation** is that the glucose is consumed by cells and that, **because the amount $V_L c_x$ of cells increases**, the total rate R_s of glucose consumption increases due to the increased amount of cells, **there are just more glucose eaters!!**

This shows that besides R_i, which is a **total rate and represents the rate performance of the whole cultivation vessel**, **a second type of rate** is relevant: the **biomass specific rate q_i, which represents the performance of the organism:**

$$q_i = \frac{R_i}{V_L c_x} = \frac{\text{mol of i produced or consumed per hour in the cultivation vessel}}{\text{Cmol of biomass present in the cultivation vessel}} \tag{15}$$

The q-rate is obtained from the total production or consumption rate (R_i, obtained from the balance for compound i) by **dividing with the total amount of biomass** ($V_L c_x$) **present in the vessel**.

These q-rates are the fundamental rates for cells, because they characterize their rate performance

The q_i-rates are influenced by:

- The properties of the cells (**genes** etc.).
- The **environment** of the cells as represented by the extracellular concentrations and conditions such as glucose, NH_4^+, O_2, CO_2, pH, T, pressure and other compounds present in the growth medium (vitamins, trace elements, ...).

2.2 q-rates and balances

We can now return to our general balances and reformulate the broth balance with the **q-rate in the conversion term**. We **replace** R_i(t) by $q_i(t)V_L(t)c_x(t)$. Note that the transfer term $T_{n,i}$ is **only present** for **volatile compounds**.

Broth balance for i (moli/h):

$$\frac{\mathrm{d}(V_L(t)c_i(t))}{\mathrm{d}t} = q_i(t)V_L(t)c_x(t) + F_{in}c_{i,in} + F_{out}(t)c_{iout}(t) \pm T_{n,i}(t) \tag{16}$$

2.2.1 Q-rates from batch experiments

The core q-rates in biological processes are μ, q_s and q_p and their calculations as function of time in batch condition is performed as follows.

2.2.2 Biomass specific growth rate μ

A special q-rate is $\mathbf{q_x}$, which is the rate of newly produced biomass per amount of biomass present in the cultivation vessel. For **historical reasons the symbol used for this rate is μ and not q_x** (which would be logical), leading to the **definition of μ.**

$$\mu(t) = \frac{R_x(t)}{V_L(t)c_x(t)} = \frac{\text{mol of x produced per hour}}{\text{Cmol of x present in the vessel}} \tag{17}$$

To obtain $\mu(t)$, according to this definition, we need:

- $V_L(t)c_x(t)$ at different times which is obtained from the **experimental** measured broth volume $V_L(t)$ and biomass concentration $c_x(t)$.
- **$R_x(t)$ which is obtained from the batch balance for biomass using these experimental measurements for $V_L(t)$ and $c_x(t)$:**

$$\frac{\mathrm{d}(V_L(t)c_x(t))}{\mathrm{d}t} = R_x(t) \tag{18}$$

Using the above **definition for $\mu(t)$,** which we use to replace $R_x(t)$ in the biomass balance, and realizing that $V_L(t)c_x(t) = N_x(t)$ gives:

$$\frac{\mathrm{d}N_x}{\mathrm{d}t} = \mu(t)N_x(t) \tag{19}$$

An essential assumption is now needed**, that μ(t) is constant** in **a small experimental time interval t_1 to t_2**, because the substrate concentration c_s changes little and can be assumed constant, **hence $\mu(t) = \mu$.** For such a **small time interval**, (because μ is constant) we can integrate the above differential equation. This integration requires separation of the variables $N_x(t)$ and t:

$$\frac{\mathrm{d}N_x(t)}{N_x(t)} = \mu \mathrm{d}t \tag{20}$$

Now, integration both left and right (from t_1 to t_2) gives:

$$\ln(N_x(t_2)) - \ln(N_x(t_1)) = \mu(t_2 - t_1) \tag{21}$$

Rearrangement then leads to the equation for μ in time interval t_1 to t_2:

$$\mu = \frac{\ln(N_x(t_2)/N_x(t_1))}{t_2 - t_1} \tag{22}$$

This **final result** allows to obtain **the -values for each time interval,** by entering for each time interval the **experimentally obtained** biomass amounts $N_x(t_2)$, $N_x(t_1)$ and $t_2 - t_1$.

Furthermore, taking exponents on both sides of equation 21 gives the **exponential growth equation**:

$$N_x(t_2) = N_x(t_1)e^{(t_2 - t_1)} \tag{23}$$

This shows that in a batch cultivation the **amount of biomass** increases **exponentially in time** (**when μ is constant**). The above result for μ, considering time intervals from t = t to t = 0 leads to a **linear equation** in μ (considering the natural logarithm). For each time point *t* the linear equation reads:

$$\ln(N_x(t_2)) - \ln(N_x(t_1)) = \mu(t_2 - t_1) \tag{24}$$

This equation can be derived for all observations t_i, leading to a classical linear regression approach:

$$\begin{aligned}
\ln\left(N_x\left(t_1\right)\right) - \ln\left(N_x\left(t_0\right)\right) &= \mu\left(t_1 - t_0\right) \\
\ln\left(N_x\left(t_2\right)\right) - \ln\left(N_x\left(t_0\right)\right) &= \mu\left(t_2 - t_0\right) \\
&\vdots \\
\ln\left(N_x\left(t_n\right)\right) - \ln\left(N_x\left(t_0\right)\right) &= \mu\left(t_n - t_0\right)
\end{aligned} \tag{25}$$

If the concentration at t=0 is not well defined, it is advised to use the redundancy of the measurements to estimate this value as well, i.e. bring it on the right hand side (RHS) (independent variables):

$$\begin{aligned}
\ln\left(N_x\left(t_1\right)\right) &= \mu\left(t_1 - t_0\right) + \ln\left(N_x\left(t_0\right)\right) \\
\ln\left(N_x\left(t_2\right)\right) &= \mu\left(t_2 - t_0\right) + \ln\left(N_x\left(t_0\right)\right) \\
&\vdots \\
\underbrace{\ln\left(N_x\left(t_n\right)\right)}_{\mathbf{y}} &= \underbrace{\mu}_{\mu}\underbrace{\left(t_n - t_0\right)}_{\mathbf{A}_a} + \underbrace{\ln\left(N_x\left(t_0\right)\right)}_{\mathbf{A}_b}
\end{aligned} \tag{26}$$

Using linear algebra, we can find a solution to this **Least Squares Linear Regression** problem. The parameters and k_0 (=ln(N_{x0})) are obtained by:

$$\begin{pmatrix} \mu \\ k_0 \end{pmatrix} = \left(\ \mathbf{A}^T\mathbf{A}\ \right)^{-1}\mathbf{A}^T \begin{pmatrix} \ln\left(N_x\left(t_0\right)\right) \\ \vdots \\ \ln\left(N_x\left(t_n\right)\right) \end{pmatrix} \quad \mathbf{A} = \begin{pmatrix} \mathbf{A}_a & \mathbf{A}_b \end{pmatrix} \tag{27}$$

The solution of this linear system will become clearer with the following example.

Example 1.2: How to obtain μ

This example is a continuation of example 1.1 from which, using $V_L(t)$ and $c_x(t)$ measurements, we calculated $N_x(t)$ (=$V_L(t)c_x(t)$).

Using these data we generate the linear equation system to estimate the growth rate and initial biomass concentration (which was also measured):

```
# batch assumptions - constant growth (substrate excess), regression to obtain mu
y = np.log( N_X )
A = np.vstack( (t, np.ones( np.size(t) ) ) ).T

# Apply linear regression to calculate mu
beta = linalg.lstsq( A, y, rcond=None )[0]
mu = beta[0]
N_X0 = np.exp(beta[1])
print( 'mu = {:6.2f}'.format( mu ))

# print( A )   # print A to understand the matrix you have generated

# fit in log and normal scale
plt.figure( figsize = (14, 6 ))
plt.subplot(1, 2, 1)
plt.plot( t, y, 'ro' )
plt.plot( t, A @ beta, 'b-' )
plt.xlabel( 'time [h]' )
```

```
plt.ylabel( 'biomass ($ln(N_X)$)')
plt.title( 'linear fit' )

tt = np.linspace( 0, np.max( t ), 100 )
plt.subplot(1, 2, 2)
plt.plot( t, N_X, 'ro', label = 'measured' )
plt.plot( tt, N_X0 * np.exp( mu*tt), 'b-', label = 'fit' )
plt.xlabel( 'time [h]' )
plt.ylabel( 'biomass [mmol/L]')
plt.title( 'BM' )
plt.legend()
plt.show()
```

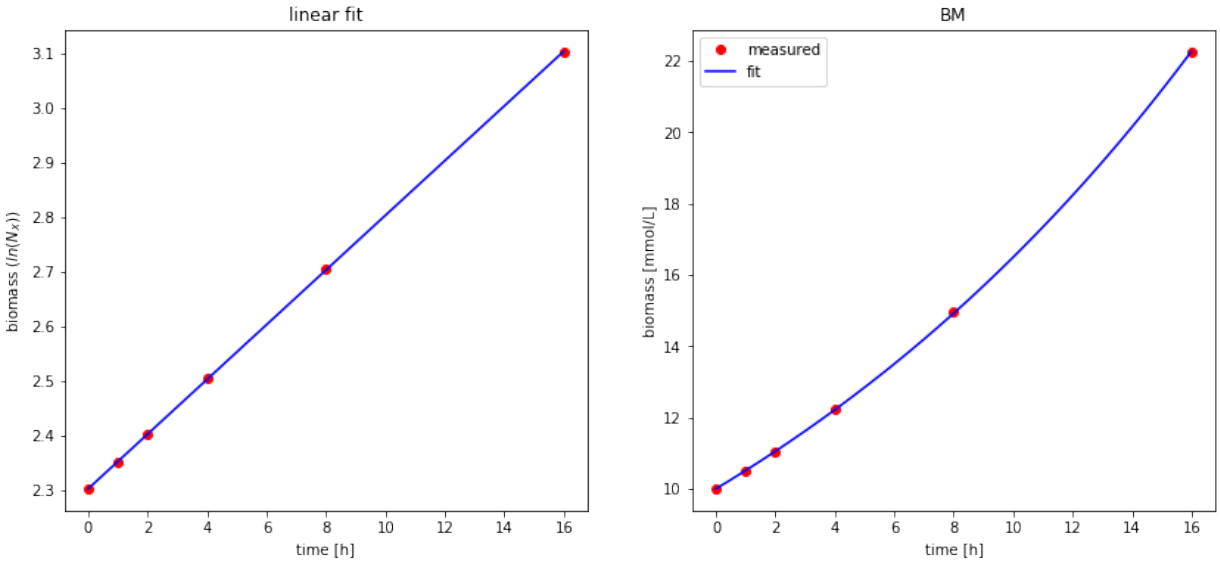

Figure 6: Left : Logarithmic plot of Nx(t) and linear regression. Right : (back) Transformation to linear scale.

2.2.3 Biomass specific substrate uptake rate q_s

According to its definition,

$$q_s = \frac{R_s(t)}{V_L(t)c_x(t)} \tag{28}$$

In this definition, $V_L(t)c_x(t)$ follows from the experimental measurements and $R_s(t)$ must be calculated from the **substrate balance**:

$$\frac{\mathrm{d}(V_L(t)c_s(t))}{\mathrm{d}t} = R_s(t) \tag{29}$$

We can eliminate $R_s(t)$ and introduce $q_s(t)$ using the q_s-definition, which leads to the substrate balance containing the desired $q_s(t)$.

$$\frac{\mathrm{d}(V_L(t)c_s(t))}{\mathrm{d}t} = q_s(t)V_L(t)c_x(t) \tag{30}$$

This differential equation is **not easy to solve** because $V_L(t)c_x(t)$ increases exponentially with time t. However, we can use the approach of **combination of balances to eliminate $V_L(t)c_x(t)$,** which removes this mathematical problem. The biomass balance gave (see before):

$$\frac{\mathrm{d}\left(V_L(t)c_x(t)\right)}{\mathrm{d}t} = \mu(t)V_L(t)c_x(t) \tag{31}$$

Elimination of $V_L(t)c_x(t)$ from **two balances** (for biomass and substrate) gives:

$$\frac{\mathrm{d}\left(V_L(t)c_s(t)\right)}{\mathrm{d}t} = \frac{q_s(t)}{\mu(t)} \cdot \frac{\mathrm{d}\left(V_L(t)c_x(t)\right)}{\mathrm{d}t} \tag{32}$$

In a small time interval (t) and $q_s(t)$ are considered constant and equal to q_s and μ. This allows integrating both sides of the above equation between t_1 and t_2.

$$\left[V_L\left(t_2\right)c_s\left(t_2\right) - V_L\left(t_1\right)c_s\left(t_1\right)\right] = \frac{q_s}{\mu}\left[V_L\left(t_2\right)c_x\left(t_2\right) - V_L\left(t_1\right)c_x\left(t_1\right)\right] \tag{33}$$

This result can be rewritten as

$$\begin{bmatrix} \text{consumed amount} \\ \text{of substrate in} \\ \text{time interval} \\ \mathrm{t_2 - t_1} \\ \mathrm{mol_S} \end{bmatrix} = \left|\frac{q_s}{\mu}\right| \begin{bmatrix} \text{produced amount} \\ \text{of biomass in} \\ \text{time interval} \\ \mathrm{t_2 - t_1} \\ \mathrm{Cmol_X} \end{bmatrix} \tag{34}$$

From this relation and experimental measurements we obtain the $\left|\frac{q_s}{\mu}\right|$ ratio by plotting consumed substrate amount (mol_S) against amount of produced biomass ($Cmol_X$). Because we know μ already, it is easy to obtain q_s. In this form, the equations are again linear in qs and can be solved applying linear regression (see example 1.3). Depending on the confidence in the initial substrate concentration, this can be set known, or estimated as well (i.e. set as independent variable in the regression):

$$
\begin{aligned}
N_s(t) - N_s(0) &= \frac{q_s}{\mu}\left(N_x(t) - N_{x,0}\right) \\
\underbrace{N_s(t)}_{\mathbf{y}} &= q_s \underbrace{\left(\frac{N_x(t) - N_{x,0}}{\mu}\right)}_{\mathbf{A}_{qs}} + \underbrace{N_s(0)}_{N_{S0}} \\
\mathbf{N}_s &= \begin{bmatrix} \frac{\mathbf{N}_x - N_{x,0}}{\mu} & \mathbf{1}^T \end{bmatrix} \begin{pmatrix} q_s \\ N_{S0} \end{pmatrix}
\end{aligned}
\tag{35}
$$

Example 1.3 Obtaining q_s from q_s/μ ratio

Using the $V_L(t)$, $c_x(t)$ and $c_s(t)$ data from Example 1.1 we can calculate $N_x(t)$ and $N_s(t)$ as function of time.

```
# batch assumptions - constant growth (substrate excess), regression to obtain mu
# assuming CX0 is error prone as well

y = N_S
A = np.vstack( ((N_X - N_X0)/mu, np.ones( np.shape(N_X) ) ) ).T
# print( A )

# Solve the linear problem using linalg.lstsq
beta = linalg.lstsq( A, y, rcond=None )[0]
qS = beta[0]
N_S0 = beta[1]
print( 'qS = {:6.2f}, N_S0 = {:6.2f}'.format( qS, N_S0 ))

# plot the results
plt.figure( figsize = (14, 6 ))

# in regression space
plt.subplot(1, 2, 1)
plt.plot( A[:,0], y, 'or', label = 'measured' )
plt.plot( A[:,0], A @ beta, 'b-', label = 'linear fit' )
plt.xlabel( '$(N_X - N_{X0})/\mu$')
plt.ylabel( '$N_S$' )
plt.legend()

# in original coordinates (t, NS)
plt.subplot(1, 2, 2)
tt = np.linspace( 0, max(t), 100 )   # generate 100 points
plt.plot( t, y, 'or', label = 'measured' )
plt.plot( tt, N_S0 + qS/mu * N_X[0] * (np.exp(mu*tt) - 1), 'b-', label = 'fit' )
plt.xlabel( '$t$')
plt.ylabel( '$N_S$' )
plt.legend()
plt.show()
```

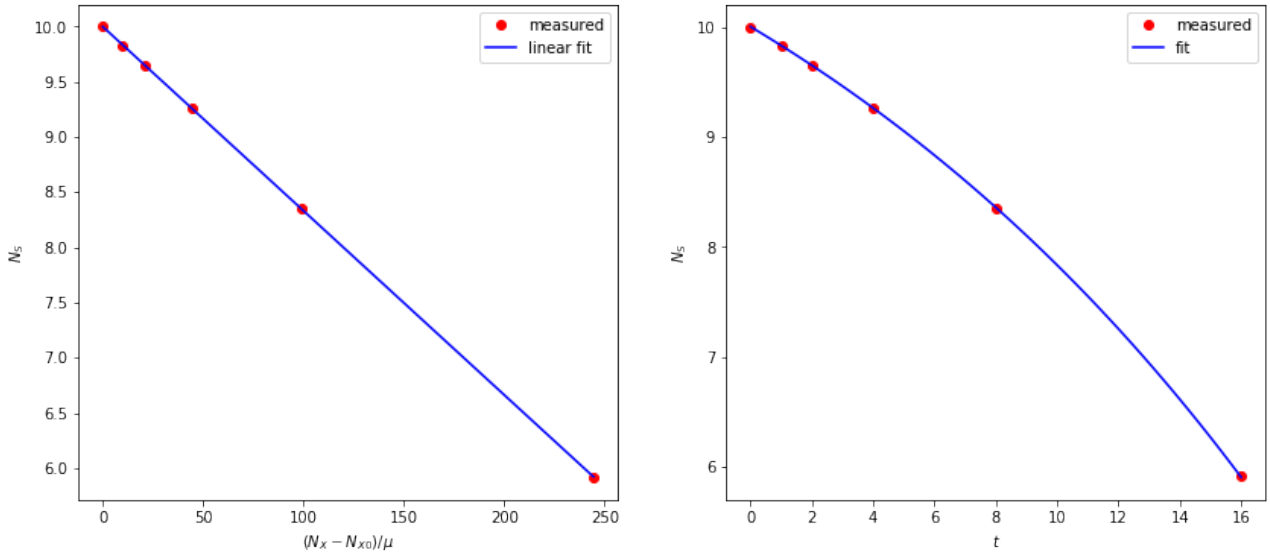

Figure 7: Consumed substrate versus produced biomass amounts. Right : Predicted timecourse using the estimated rates.

Note, integration of $\frac{\mathrm{d}N_s}{\mathrm{d}t} = q_s N_x$ results in $N_s = N_s, 0 + \frac{q_s}{\mu} N_{x,0}\,(e^{\mu t} - 1)$.

2.2.4 Biomass specific product formation rate q_p

q_p can be calculated in the same way as q_s. One needs **to combine the batch balances for product and biomass** leading to a relation for the ratio q_p/μ. From the measurements $V_L(t)$, $c_x(t)$, $c_p(t)$ one calculates at each time point $V_L(t)c_p(t)$ and $V_L(t)\,c_x(t)$. One obtains the q_p/μ ratio for each time interval using:

$$\frac{q_p}{\mu} = \frac{V_L\,(t_2)\,c_p\,(t_2) - V_L\,(t_1)\,c_p\,(t_1)}{V_L\,(t_2)\,c_x\,(t_2) - V_L\,(t_1)\,c_x\,(t_1)} = \frac{\text{mol produced product between } \mathrm{t_2} \text{ and } \mathrm{t_1}}{\text{mol produced biomass between } \mathrm{t_2} \text{ and } \mathrm{t_1}} \tag{36}$$

Knowing μ (see above) and the obtained q_p/μ ratio immediately gives q_p.

2.3 The volume specific rate r_i

Until now, we have introduced two types of rates: R_i (total rate of compound i, in mol(i)/h), and q_i (the biomass specific rate in mol(i)/(Cmol(x) $\times$ h).

In economic considerations, the investment of equipment is related to the total broth volume (V_L) in the cultivation vessels. From this point of view, it is important to have information of the **volume specific rate which is named** R_i.

$$r_i = \frac{R_i}{V_L} \qquad \text{in} \frac{mol(i)/h}{m^3 \text{of broth}} \tag{37}$$

The definition for q_i and R_i shows that

$$r_i = q_i c_x \tag{38}$$

Hence the volume specific rate r_i and the biomass specific rate q_i are coupled through the **biomass concentration c_x.**

2.4 Conclusion on obtaining μ, q_s, q_p from batch experiments

One uses the proper measurements (on time dependent concentrations, volumes etc.) to calculate the **amounts** $N_i(t) = V_L(t)c_i(t)$ as function of time (for i = X, S, P).

Subsequently one uses the **biomass balance** to obtain μ from the $N_x(t)$ data.

Using combinations **of balances** (for substrate and biomass respectively product and biomass) one obtains the relations which give the q_s/μ and q_p/μ ratios from **linear plots** of produced or consumed amounts, (substrate, biomass and product, biomass). The obtained q_s/μ, and q_p/μ ratios give, knowing μ, the values for q_s and q_p.

2.5 Behavior of the batch process

We have seen that in a batch process the following applies

- $\mu(=\mu_{max})$, $q_s(=q_{s,max})$ and $q_p(=q_{p,batch})$ are independent of time in a batch. **Hence q-rates are constant**.
- The biomass amount present in the whole fermentor $N_x(t)$ **increases exponentially in time**

  $N_x(t) = N_x(0)$ exp [μ_{max} (t-0)]due to a constant growth rate μ_{max}.
- The amount of produced biomass follows as $N_x(t) - N_x(0)$, which is known for known $N_x(0)$ (inoculum) and $N_x(t)$ which follows from the above exponential biomass relation.
- The conversion rates in the total batch $R_i(t) = q_i N_x(t)$ **increase exponentially in time.**
- The ratio q_s/μ is constant in time ($q_{s,max}/\mu_{max}$) and known. From the ratio we can calculate:

$$N_s(t) - N_s(0) = \frac{q_s}{\mu}\left(N_x(t) - N_x(0)\right) \tag{39}$$

  In a similar way, using the constant q_p/μ ratio, one can obtain $N_p(t)$ from $N_x(t) - N_x(0)$

Fig.8 shows the typical time profiles for amounts $N_i(t)$, $R_i(t)$ and q_i in batch.

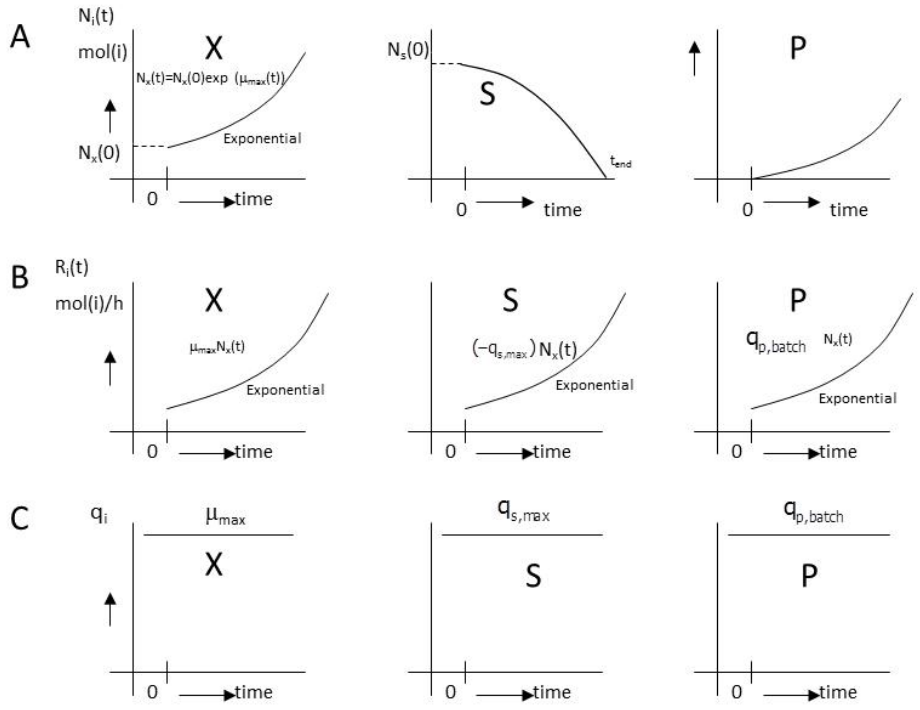

Figure 8: Dynamics in a batch process of (A) : $N_i(= V_L(t)c_i(t))$, (B) : R_i and (C) : q_i for biomass (x), substrate (s), product (p).

2.6 Bananas: pitfalls in obtaining q-rates from batch experiments

2.6.1 Introduction

Because the performance of organisms is quantified by their q-rates it is important to be aware of pitfalls which lead to wrong q-values. We can distinguish two categories, being experimental pitfalls and calculation pitfalls. Because such pitfalls are similar to unprepared people falling down due to slippery banana peels, we call this “bananas".

2.6.2 Experimental banana: how to obtain an adapted inoculum

The presented theoretical description of batch experiments **can only be observed in practice (as shown in Fig.8) using an adapted inoculum**. In particular, it is assumed that the used inoculum biomass immediately does grow at μ^{max}. This is **only assured** if the inoculum was taken from a culture **growing exponentially** in batch at the **same conditions** (pH, T, nutrients composition, **substrate excess**) as in the intended new batch. This is called an **adapted inoculum**.

2.6.3 Problems in obtaining an adapted inoculum

In practice this is difficult to realize and therefore one often inoculates with non-adapted biomass. This happens

- When the batch inoculum cultivation is already finished before one takes the inoculum. This means that the inoculum biomass has been without substrate for a certain time period, which leads to changes in the organism (aimed e.g. at survival).
- When the batch cultivation is performed in shake flasks, then often O_2-supply becomes limiting which changes the condition from substrate excess (and excess of all other nutrients) to a condition of O_2-limitation. This leads to changes in the organism.

- Again, in shake flasks, there is no pH-control and there may be a pH-change in the inoculum cultivation. This pH of the inoculum culture might be very different from the pH intended to be used in the batch experiment for which the inoculum is cultivated. Again, the organism properties will change due to the different pH.
- Frequently, one uses a “rich" nutrient medium for the inoculum culture, which is different from the batch medium!!
- Frequently, one harvests inoculum biomass from e.g. shake flasks (which are practical) but subjects the biomass, before adding it to the fermentor e.g. to washing (to remove nutrients from the inoculum growth medium) and to storage in the fridge (because the batch fermentor is not ready yet!). This causes osmotic and temperature shocks for the organism, leading to genetically based changes.

Clearly any deviation of the intended batch fermentor condition (inoculum shake flask nutrient composition $\neq$ fermentor batch nutrient composition, inoculum washing or storage, absence of substrate excess, exposure to O_2-limitation, deviating pH and temperature) will affect the kinetic properties of the inoculum biomass, which is then a **non-adapted inoculum**.

This situation leads to a so called **lag phase** in the batch fermentor.

This is a time period after inoculation where the organism, coming from the shake flask condition, adapts to the conditions in the batch fermentor culture.

During this lag-period the exponential behavior and constant q-values as predicted by the model (Fig.8) do **not apply**.

2.6.4 How to prepare an adapted inoculum

The best method is to execute a **repeated batch experiment** in a fermentor

- In a fermentor one can **control** Temperature, pH, dissolved O_2, the CO_2-concentration etc. This control is **not** possible in a shake flask (which is a popular method to prepare inoculum because it is so simple)
- In a **repeated batch,** one starts with a non-adapted inoculum prepared in shake flasks. The first batch will show a lag-phase. Before the end of the first batch one removes 90-95$\%$ of the broth and adds quickly fresh batch cultivation medium. The remaining 5-10$\%$ broth contains then the inoculum biomass for the second batch. This can be repeated several times and one will observe that from the 2nd or 3rd batch one does **not** observe a lag-phase anymore!!

To calculate the active fraction of inoculum, one can perform a series of experiments using different inoculum densities (see assignment A1 from 2012).

2.6.5 How to obtain wrong q-rates

In the previous sections we have introduced the **total rate** R_i and **the biomass specific rate q_i** and it has been shown how these can be obtained from measurements and proper balances, using the batch experiment as an example. Below additional examples for these **balance calculations** will be presented, where each example contains a pitfall (**banana**) which causes wrong rates, leading to a **message.**

Example 1.4: Wrong balances

Many students write the following batch balances

Substrate balance:

$$\frac{\mathrm{d}(V_L c_s)}{\mathrm{d}t} = q_s V_L c_x \tag{40}$$

Biomass balance

$$\frac{\mathrm{d}(V_L c_x)}{\mathrm{d}t} = \mu V_L c_x \tag{41}$$

Product balance

$$\frac{\mathrm{d}(V_L c_p)}{\mathrm{d}t} = q_p V_L c_x \tag{42}$$

Task: Find the banana!!

Example 1.5

Mistakes in evaluating the increase (or decrease) term of the balance, **do not neglect the change in V_L and use the correct V_L as defined by your concentration measurement**

A correct calculation of R_x and R_s under batch conditions depends on a correct evaluation of the increase (or decrease) term in the balances using experimental data. **Evaluation of the increase (or decrease) term is not trivial**.This term can be expanded as:

$$\frac{\mathrm{d}\left(V_L c_i\right)}{\mathrm{d}t} = V_L \frac{\mathrm{d}c_i}{\mathrm{d}t} + c_i \frac{\mathrm{d}V_L}{\mathrm{d}t} \tag{43}$$

This shows that the increase (or decrease) in the time or in the amount ($V_L c_i$) of molecules of compound i present in the cultivation vessel has ***two* contributions**: concentration ($V_L \frac{\mathrm{d}c_i}{\mathrm{d}t}$), and volume ($c_i \frac{\mathrm{d}V_L}{\mathrm{d}t}$).

A **first** often-made **mistake** is that one looks **only** at concentrations. For example, when c_i is measured to be constant ($dc_i/dt = 0$) one concludes that an increase or decrease of i is absent and therefore there is no reaction, hence $R_i = 0$. **This is only true when also the volume is constant.**

Clearly the change in the amount $V_L c_i$ ($= N_i$), **and not the change in c_i**, is always needed, which requires using the measured **changes in time** of in **both V_L and c_i**.

A **second** mistake, relates to the **volume that should be used**. This is *not* trivial. Suppose one obtains the concentration c_i by taking a broth sample and one removes the biomass to obtain **a stable supernatant** (because the biomass in the sample e.g. consumes substrate and produces compounds, which changes concentrations). Therefore, one does analyze the concentration of the compound of interest in the supernatant and one also measures the volume fraction of the supernatant in the total broth sample volume. The supernatant volume is smaller than the broth volume due to the **volume of wet biomass**. The total amount in the vessel should now be calculated by multiplying this c_i-value (mol of i per m^3 supernatant) with **the volume of supernatant in the vessel** (supernatant volume fraction in the sample times total broth volume). This is **not the same** as using the volume of broth in the vessel (which is supernatant + wet biomass). The difference depends on the biomass volume, but can easily be 5-20$\%$!! However, this approach **assumes** that the compound i is **only present** in the supernatant and is not present inside the biomass. **This is not always so.** It may well be that there is a significant product concentration present inside the biomass. Only when there is complete secretion of the product by active export over the cell membrane it can be expected that its concentration inside the biomass is negligibly low. In all other situations, the amount of product in the biomass must be quantified separately.

2.7 Take-home messages (q-rates, balances and batch experiments)

The **biomass specific** q_i-rates define the performance of the organisms. The **core q-rates are:** μ, q_s, q_p.

These q-rates must be **calculated** from **balances** combined with measurements during experiments where the cell, tissue or microorganism is consuming, growing and producing (in a Petri-dish, flow cell, fermentor, shake-flask, etc.). A q-rate must be calculated because it **cannot be measured by a sensor**.

Key to these calculations are to make proper system definitions and to **correctly** formulate **the balances** for the defined system.

These balances:

- are set-up **for each compound** i
- require defined **system boundaries**

- have all the **same format** (increase or decrease of the amount of the compound within the boundary = prod/cons within the boundary + transport over the boundary)
- have the **unit** mol_i/h
- use a general **set of symbols**
- force to reflect and decide on **mechanisms** (**conversion** and/or **transport**) which act **simultaneously** on the compound for which the balance is made. **Each mechanism** leads to **its own term on the balance**
- show which **measurements** must be performed to calculate q_i from its balance and these measurements

In **batch** (**excess nutrient**), using an **adapted inoculum:**

- q-values are **constant in time**
- the biomass amount ($V_L c_x$) **increases exponentially in time**
- all total **rates** $R_i = q_i V_L c_x$ increase **exponentially in time**
- the stoichiometry reflected in **constant** q_i/μ ratio's q_s/μ, q_p/μ, is **constant in time**

However even for the simple batch experiments, serious **banana's** exist in the correct experimental design (adapted inoculum needed) and correct evaluation of q-rates due to errors in setting up the balances.

3 Recovery, Reconciliation and Hidden Processes

3.1 Introduction

Any quantitative evaluation of microbial processes requires a good basis for data quality. One possibility to check data quality is the recovery of elements. Especially,

- The measurements should cover all relevant processes
- The incoming material should be found back in the products (biomass, CO_2, products, etc)
- All elements should be found back (carbon, nitrogen, charge as well as degree of reduction)

In the list, oxygen and hydrogen are not mentioned as elements to be recovered. The simple reason is the production of water by microorganisms. This is a very small amount compared to the water in the system and cannot be quantified. Good news is that water has a degree of reduction of zero and therefore will not appear in the degree of reduction balance.

In case that the recovery is incomplete, i.e. not all carbon, degree of reduction or nitrogen are found back in the products:

- there can be processes that are not monitored, for example a byproduct that is not measured (in the liquid or gas phase)
- there are issues with the calculation (in- outflows in the liquid / gas)
- there are issues with the measurements

Based on the observed gaps, hypotheses can be derived. Especially useful is to analyze the **degree of reduction to carbon ratio**. A classic example for a missing process is biomass lysis. Biomass has a degree of reduction in the range of 4.2-4.6 electrons/carbon. If the ratio of the gap is in that range, measurements for lysed biomass should be considered.

To identify whether additional carbon-containing compounds are present in the liquid phase, the total organic carbon can be measured (the so-called TOC measurements). These measurements can be performed using broth and filtrated broth to additionally identify whether the yet unknown compound is inside or outside the biomass.

Furthermore, elemental conservation leads to redundancies in the measurements – this redundancy can be exploited to improve the rate estimation. This approach is called data reconciliation – i.e. the use of additional information to obtain consistent rates that fulfill the elemental balances.

3.2 Element recovery

Elements cannot be converted by a microbial process (this only happens in for example nuclear reactors or nuclear fusion). Therefore, the elements (C,N,O,H) have to be recovered in the process products. How much of the substrate elements are found back is quantified in the so called recovery (R_E):

$$R_E = \frac{\sum \text{elements in products}}{\sum \text{elements in substrates}} \tag{44}$$

Please note that the recovery can be higher than 100$\%$ (for example in case of an overestimation of the product concentrations / calculation errors). Furthermore, the recovery is based on flows – we can define the recovery of the bioprocess, or based on biomass specific rates (specific recovery). It is advised to start with bioreactor recoveries (defining the reactor as the system to study) as opposed to biomass recoveries. Biomass normalization can be biased in the case of changes in the biomass composition (for example intracellular product accumulation). The gap in the element balance is defined as:

$$\Delta_E = \sum \text{elements in product rates} - \sum \text{elements in substrate rates} \tag{45}$$

These calculations can be easily performed using excel. Here, we will use Python – not because it appears a bit more complicated, but because we will need the elemental balances in later calculation steps (reconciliation) in matrix format. For the moment we will focus on the elemental balances of carbon, nitrogen and degree of reduction. With these three balances, the charge balance as well as oxygen and hydrogen are also closed (given that there are no metal oxidation steps or other alternative electron sources/sinks). In view of the upcoming steps, the elemental balances will be formulated in matrix notation. Let's have a look at an example process – *S. cerevisiae* aerobic conditions, with aerobic fermentation (high substrate uptake in batch). The processes to take into account are:

- Substrate uptake (glucose feeding):rate R_S
- Biomass formation rate R_X
- Ethanol formationrate R_{EtOH}
- Carbon dioxide formationrate R_{CO2}
- Oxygen uptakerate R_{O2}

Carbon containing are glucose, biomass, ethanol and carbon dioxide. For the carbon-balance we need to take the amount of carbon per molecule into account (i=substrates, j=products). It is assumed that the biomass rate is derived from dry-weight measurements and converted to Cmol/g_{CDW}/h (using $M_W = 24.6$ g/Cmol):

$$\begin{aligned}
&R_{C,\text{in}} = \sum \#\text{ carbons }_j \cdot \text{ rate }_j, R_{C,\text{out}} = \sum \#\text{ carbons }_i \cdot \text{ rate }_i, \quad R_C = \frac{R_{C,\text{out}}}{R_{C,\text{in}}} \\
&R_{C,\text{in}} = 6R_S \quad R_{C,\text{out}} = 2R_{EtOH} + 1R_{CO2} + 1R_X \\
&R_{C,\text{out}} = \mathbf{E}_{\text{out}}\,\mathbf{R}_{\text{out}} \quad \mathbf{E}_{\text{out}} = \begin{pmatrix} 2 & 1 & 1 \end{pmatrix} \quad \mathbf{R}_{\text{out}} = \begin{pmatrix} R_{EtOH} \\ R_{CO2} \\ R_X \end{pmatrix} \\
&R_{C,\text{in}} = \mathbf{E}_{in}\mathbf{R}_{\text{in}} \quad \mathbf{E}_{\text{in}} = (6) \quad \mathbf{R}_{\text{in}} = (R_S)
\end{aligned} \tag{46}$$

The unit of the rates $R_{C,in}$ and $R_{C,out}$ will be in Cmol/h (resp. other time unit). The recovery (fraction $R_{C,out}/R_{C,in}$) will be expressed in $\%$. Reasonable recoveries are in the range of 95-105$\%$ (depending also on the accuracy of the measurements).

Similarly, the recovery of the degree of reduction can be formulated. Now, we have to take into account the amounts of electrons per molecule. Clearly, all compounds that contain electrons need to be taken into account (substrate, biomass, oxygen, ethanol):

$$
\begin{aligned}
&R_{e-,\,\text{in}} = \sum \#\,\text{electrons}_{\,j} \cdot \text{ rate}_{\,j}, R_{e-,\,\text{out}} = \sum \#\,\text{electrons}_{\,i} \cdot \text{ rate}_{\,i}, \quad R_C = \frac{R_{C,\,\text{out}}}{R_{C,\,\text{out}}} \\
&R_{e-,\,\text{in}} = 24R_S - 4R_{O2} \quad R_{e-,\,\text{out}} = +12R_{EtOH} + 0R_{CO2} + 4.2R_X \\
&R_{e-,\,\text{out}} = \mathbf{E}_{\text{out}}\,\mathbf{R}_{\text{out}} \quad \mathbf{E}_{\text{out}} = \begin{pmatrix} 12 & 0 & 4.2 \end{pmatrix} \quad \mathbf{R}_{\text{out}} = \begin{pmatrix} R_{EtOH} \\ R_{CO2} \\ R_X \end{pmatrix} \\
&R_{e-,in} = \mathbf{E}_{in}\mathbf{R}_{in} \quad \mathbf{E}_{in} = \begin{pmatrix} 24 & -4 \end{pmatrix} \quad \mathbf{R}_{in} = \begin{pmatrix} R_S \\ R_{O2} \end{pmatrix}
\end{aligned}
\tag{47}
$$

These equations can also be used to calculate the gaps, i.e. the rate of missing carbon / electrons:

$$
\Delta_C = R_{C,\,\text{in}} - R_{C,\,\text{out}} = \mathbf{E}_{\text{in}}\,\mathbf{R}_{\text{in}} - \mathbf{E}_{\text{out}}\,\mathbf{R}_{\text{out}} \tag{48}
$$

In the case of the 'ideal' experiment, the gap will be 0. As this has to be true for a process, we will later use this as an equation (). To have a compact notation, we can bundle the in and out vectors (only difference is the sign):

$$
\begin{aligned}
\Delta_C &= R_{C,\,\text{in}} - R_{C,\,\text{out}} = \mathbf{E}_{\text{in}}\,\mathbf{R}_{\text{in}} - \mathbf{E}_{\text{out}}\,\mathbf{R}_{\text{out}} \\
&= \mathbf{ER} \qquad \mathbf{E} = \begin{pmatrix} \mathbf{E}_{\text{in}} & \mathbf{E}_{\text{out}} \end{pmatrix} \quad \mathbf{R} = \begin{pmatrix} \mathbf{R}_{\text{in}} \\ -\mathbf{R}_{\text{out}} \end{pmatrix}
\end{aligned}
\tag{49}
$$

Similarly, the equations for the different elements can be bundled. For the S. cerevisiae case, we can define a vector with all rates and a matching E matrix, containing in the first row the C-balance, in the second row the electron balance:

$$
\mathbf{R} = \begin{pmatrix} \mathbf{R}_{\text{in}} \\ -\mathbf{R}_{\text{out}} \end{pmatrix} = \begin{pmatrix} R_S \\ R_{O2} \\ -R_{EtOH} \\ -R_{CO2} \\ -R_X \end{pmatrix} \quad \mathbf{E} = \begin{pmatrix} 6 & 0 & 2 & 1 & 1 \\ 24 & -4 & 12 & 0 & 4.2 \end{pmatrix} \tag{50}
$$

$$
\Delta_E = \mathbf{ER}(= 0)
$$

Work with the example

You can work with this example in Python (Jupyter notebook format) on Brightspace (Sc_process). For this, you also need to download the excel file: `Data_Batch`. We will solve this exercise in class.

Example process data – Ethanol evaporation?

Experiments with *S. cerevisiae* in chemostat are performed at different dilution rates to identify the critical dilution rate (when ethanol production starts) and the yield parameters. The experiments are performed in 2L bioreactors, with 1L broth volume, an aeration rate of 0.5 L/min (specific volume 22.4 L/mol, 79% N_2, 21% O_2). The molar weight of biomass is 24.6 g/Cmol, the nitrogen content is 0.2 N/C, the degree of reduction 4.2 e/C. The following experimental results are obtained (Table1). The medium contains 150 mmol/L glucose, and 130 mmol/L ammonium.

Dilution rate	BM (g/L)	EtOH (mmol/L)	Glc (mmol/L)	NH4 (mmol/L)	TOC (broth)	TOC (filt)	CO2 (%)	O2 (%)
0.05	13.23	0.0	0.02	22.45	538	0	1.349	19.730
0.1	13.23	0.0	0.03	22.46	538	0	2.695	18.464
0.15	13.23	0.0	0.05	22.48	538	0	4.039	17.200
0.25	13.22	0.0	0.09	22.51	538	1	6.716	14.680
0.28	11.39	28.2	0.13	37.39	520	57	7.206	14.877
0.3	7.61	88.6	0.24	68.16	488	179	7.031	16.502
0.33	5.34	127.9	0.50	86.59	476	259	7.260	17.386
0.35	4.56	142.8	0.80	92.96	476	290	7.506	17.653

Table 1: Experimental results at different dilution rates.

Tasks

The data quality should be evaluated to test whether all relevant process parameters have been measured. Therefore, calculate the recovery in Carbon, Nitrogen and degree of reduction.

Test the hypothesis of ethanol evaporation. If evaporation was the mechanism, all missing carbon is expected in the gas-phase (check with the respective TOC measurements).

Evaporation commonly is a simple linear function of the concentration (all other parameters like temperature, aeration rate, pressure are kept constant) – plot the missing carbon as a function of the ethanol concentration.

Solution:

Calculate the recovery in Carbon, Nitrogen and degree of reduction.

Different options here. We focus on the conversion process rates (thus the transport rates are not included explicitly, but used to calculate the conversion rates). The process is described with the following rates (in mmol/L/h)

- Substrate uptake rate R_S
- Oxygen uptake rate R_{O2}
- Biomass synthesis rate R_X
- Ethanol synthesis rate R_{EtOH}
- Carbon dioxide rate R_{CO2}

Step 1: Read data and plot:

```
    # import relevant libraries, plot data

import numpy as np
import numpy.linalg as linalg
import pandas as pd
import os
import matplotlib.pyplot as plt

# Get the data from the excel sheet
cwd  = os.getcwd()
path = cwd + '\\Data_Sc_process.xlsx'
df   = pd.read_excel(path, header=[0] )
data = np.array(df)

data_header = list( df.columns )
```

```
D  = data[:,0]

# generate data plots
plt.figure( figsize = (14, 10 ))
c = 1
for i in [1,4,2,3,7,8]:
    plt.subplot(2, 3, c)
    plt.plot( D, data[:,i], 'ko:' )
    plt.xlabel( 'dilution rate [1/h]' )
    plt.ylabel( data_header[i] )
    plt.title( data_header[i] )
    plt.grid()
    c = c + 1

plt.savefig( "Sc_data.png" )
plt.show()
```

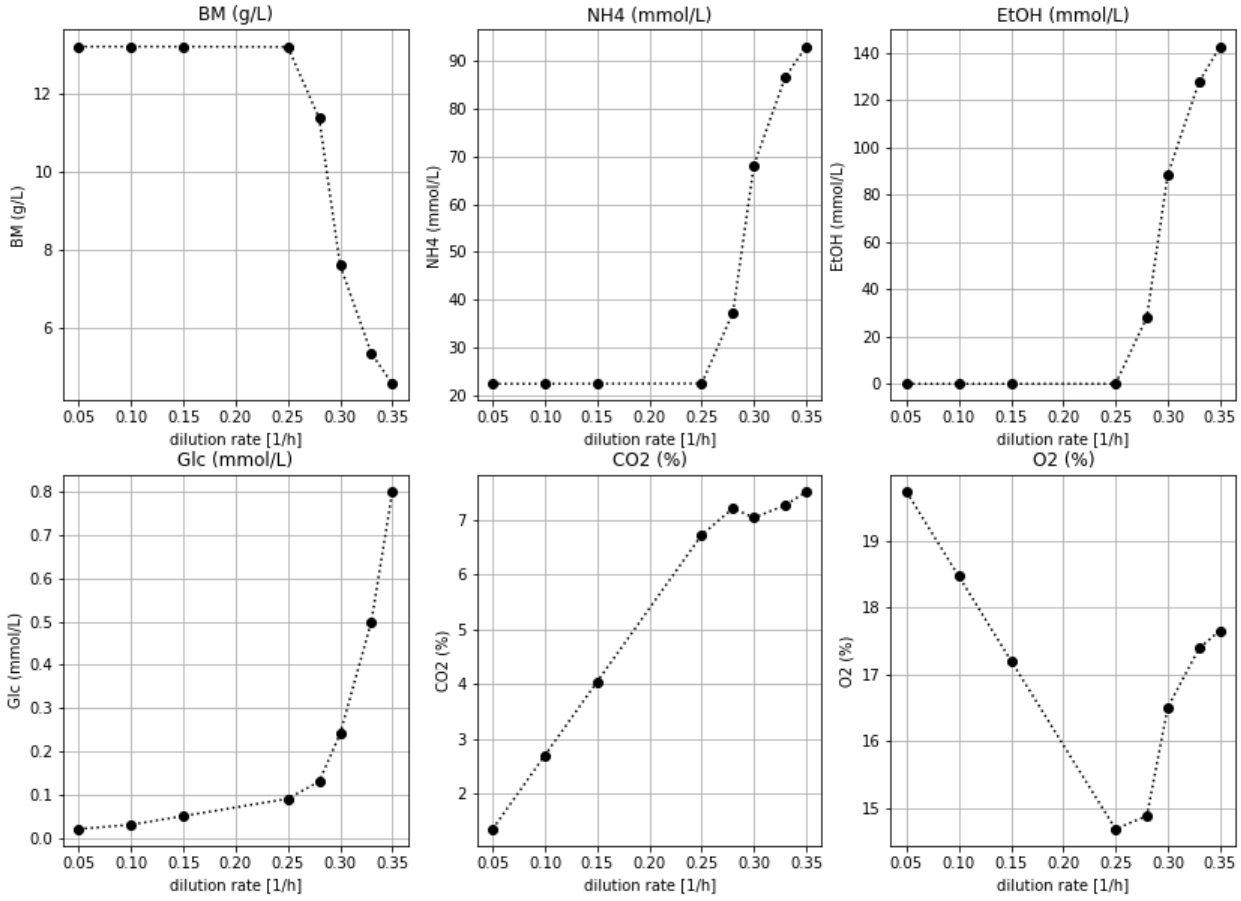

Figure 9: Observed concentration as a function of the applied dilution rate.

Step 2: calculate rates (from balances), and take care of units!

Here we use mol/L/h, all table data is converted to these (see code)

```
# known parameters
M_X = 24.6          # g/mol
N_X = 0.2
e_X = 4.2

# calculate bioreactor specific rates
D  = data[:,0]
gas_in = 0.5/22.4 * 60      # L/min ->mol/h
gas_out = 0.79 / (1-(data[:,7]+data[:,8])/100) * gas_in

# uptake rates
RS = 150e-3 * D - data[:,3]*1e-3 * D              # uptake in mol/h/L
RO2 = gas_in * 0.21 - gas_out * data[:,8]/100     # O2 in mol/h/L
RNH4 = 130e-3 * D - D * data[:,4]*1e-3

# production rates
RX    = D * data[:,1] / M_X  # biomass synthesis rate in Cmol/h/L
RCO2  = gas_out * data[:,7]/100   # CO2 in mol/h/L
REtOH = D * data[:,2]*1e-3       # ethanol in mol/l/h
```

Step 3: Recoveries

1. Prepare element matrix (in excel)

Element	BM	EtOH	Glc	NH4	TOC (broth)	TOC (filt)	CO2 (%)	O2 (%)
C	1	2	6	0	1	1	1	0
N	0.2	0	0	1	0	0	0	0
e	4.2	12	24	0	0	0	0	-4

2. Lump in- and out- rates (R_{in}, R_{out})
3. Perform calculations (E dot R)

```
# bundle in one vector (in rows: the different rates RS,RO2,RNH4, in columns the different dilution rates)
Rout = np.vstack( (RX,REtOH,RCO2) )
Rin = np.vstack( (RS,RO2,RNH4) )

# get element composition (from excel sheet, same order as measurements)
df = pd.read_excel(path, header=[0], sheet_name="C_N_e", index_col=[0] )
Emat = np.array( df )

# select columns for Ein and Eout:
Ein = Emat[ :,[2,7,3]]
Eout = Emat[ :, [0,1,6]]

gap = np.zeros( (3,D.size) )       # storage for the gaps in C, N, and e
rec = np.zeros( (3,D.size) )       # storage for the gaps in C, N, and e

print( ' D   C (%)  N (%)   e(%)' )
print( '-'*4, '-'*6,  '-'*6,  '-'*6      )

for iD in range( D.size ):
    igap = Eout @ Rout[:,iD] - Ein @ Rin[:,iD]
    irec = 1 + igap / (Ein @ Rin[:,iD])

    gap[:,iD] = igap
    rec[:,iD] = irec

    print( '{:4.2f} {:6.2f} {:6.2f} {:6.2f} '.format( D[iD], irec[0] * 100, irec[1] * 100 , irec[2] * 100  ) )
```

Dilution (1/h)	carbon	nitrogen	degree of reduction	Gap e/C
0.05	100	100	100	
0.1	100	100	99.9	
0.15	100	100	99.9	
0.25	100	100	99.9	
0.28	96.6	100	92.6	6
0.3	90.2	100.1	81.8	6
0.33	87.2	100	78	6
0.35	86.5	100.1	77.4	6

Table 2: Calculated recoveries for the different dilution rates.

From the results, it is clearly observed that as long as there is no ethanol formation (D<=0.25), the data is consistent, with fully closing balances. The always closed nitrogen balance suggests that the biomass measurements and also composition is correct and stable for the different dilution rates. The recovery of carbon decreases with increasing ethanol formation (see also Fig.10). The increasing missing carbon is linearly correlated with the ethanol concentration, suggesting that this is a missing process.

To check for a putative forgotten/hidden process, we look into the gaps and the ratio (last column). For the dilution rates with significant gap, there is a stable e/C ratio of 6.0. This ratio is the same as the e/C ratio for ethanol, thus there could be a gap in the balances due to ethanol evaporation. If this is the case, there should also be a relation between the broth EtOH concentration and the gap (evaporation is approximately proportional to the concentration in water). To check this hypothesis, the gap is plotted as a function of the EtOH broth concentration.

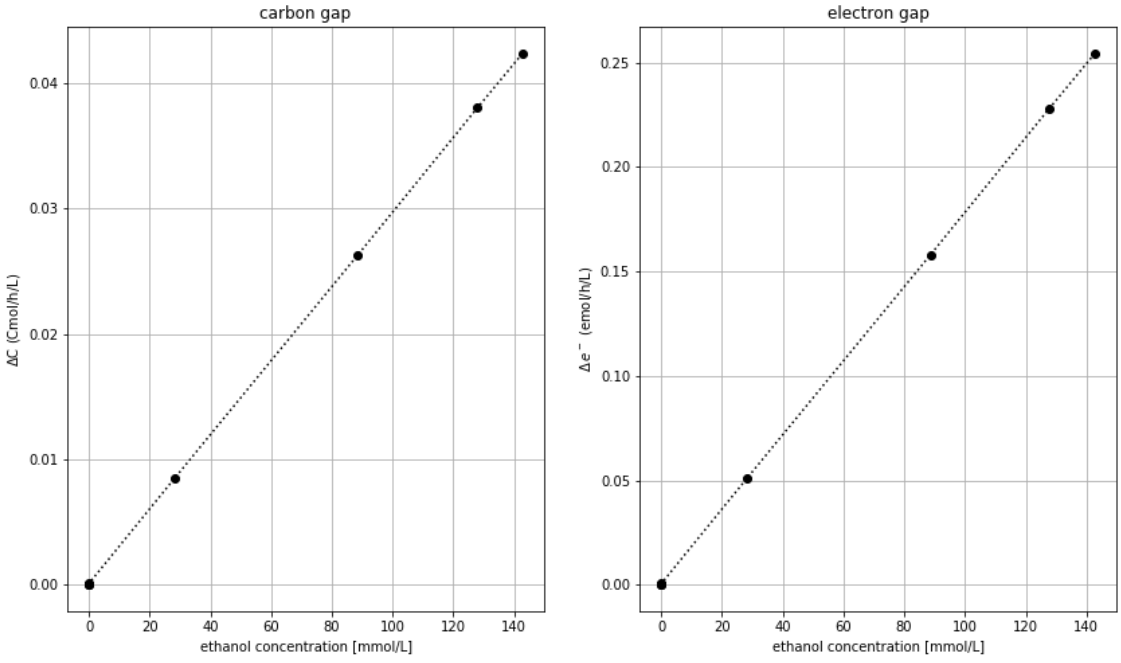

Figure 10: Left : The observed gap in carbon as a function of the ethanol concentration. A clear linear correlation in observed, suggesting that ethanol evaporation is taking place. Right : Gap in degree of reduction. The slope ratio is 6 (e-/C) -which matches the e-/C ratio of ethanol.

3.3 Reconciliation

In the previous section we observed that there can be gap's in the elemental recoveries – which violates the law of element conservation. In case of small gaps, one can argue that those gaps originate from measurement inaccuracies, which is not unrealistic given the challenges to measure compounds in a very complex background.

The aim of reconciliation is to estimate rates that fulfill the element conservation and are as close as possible to the observations. This is a classical optimization problem under linear boundary (equality) conditions. We can formulate this optimization problem as minimization of the distance between observation and estimation under the condition that the elemental balances (**ER**=**0**) are fulfilled:

$$\widehat{R} = \underset{\mathbf{ER}=\mathbf{0}}{\arg\min} \left(\| R_{\text{meas}} - R \| \right) \tag{51}$$

The linear equality constraint complicates the solution approach – the mathematically correct solution can be obtained using Lagrangian multipliers, or re-definition as a determined system (reduced set of variables) and "standard" regression – and probably there are more paths to the solution.

To keep it simple here, we use the Lagrangian approach without going into the math. Simply speaking, we solve two sets of equations:

1. The element conservations, $\mathbf{ER} = 0$ (linear, equality constraint)
2. The measurement equations $\mathbf{R_m} = \mathbf{MR} + \epsilon$ (with unknown error ϵ) – the "optimization equations, i.e. e should be minimized)

For the running example, we have observations on the rates R_S, R_{O2}, R_{NH4}, R_{EtOH}, R_{CO2} and R_X. These translate to equations:

$$\left.\begin{array}{l} R_{m,\mathrm{O2}} = R_{O2} + \varepsilon_{O2} \\ R_{m,S} = R_S + \varepsilon_S \\ R_{m.\mathrm{EIOH}} = R_{\mathrm{EIOH}} + \varepsilon_{\mathrm{EIOH}} \\ R_{m,X} = R_X + \varepsilon_X \\ R_{m,\mathrm{CO2}} = R_{\mathrm{CO2}} + \varepsilon_{\mathrm{CO2}} \end{array}\right\} \mathbf{R}_m = \mathbf{MR} + \varepsilon \tag{52}$$

Exercise: Apply reconciliation to the previous example dataset

For the low dilution rates (i.e. D <= 0.25 1/h) no changes are expected, element conservation is basically already fulfilled (99.9-100%). For the higher dilution rates, changes are to be expected, most consistently it would be to observe a higher ethanol production rate (as the measured observation misses the evaporated fraction), but we will come to a more consistent approach later.

Step 1: Define the process rates, measurements and respective matrices:

For **R**, we take a sorting R = [-Rin, Rout], and the **E** matrix constructed from E = [Ein, Eout]:

The measurements directly correspond to process rates, M becomes an eye matrix.

$$\mathbf{R} = \begin{pmatrix} -R_S \\ -R_{O2} \\ -R_{NH4} \\ R_X \\ R_{EtOH} \\ R_{CO2} \end{pmatrix} \quad \mathbf{R}_m = \begin{pmatrix} -R_{m,S} \\ -R_{m,O2} \\ -R_{m,NH} \\ R_{m,X} \\ R_{m,EtOH} \\ R_{m,CO2} \end{pmatrix} \mathbf{M}_m = \begin{pmatrix} 1 & & & & & \\ & 1 & & & & \\ & & 1 & & & \\ & & & 1 & & \\ & & & & 1 & \\ & & & & & 1 \end{pmatrix} \tag{53}$$

```
# measurement vector:
Rm = np.vstack( (-Rin, Rout) )
E = np.hstack( (Ein, Eout ) )

print( Rm )
print( E )

M = np.eye( 6 )
Rr = np.zeros( Rm.shape )
```

Step 2: Perform reconciliation for each dilution rate

```
def lagrange_solve( E, M, Rm ):

    L1 = np.hstack( (M.T @ M, E.T) )
    L2 = np.hstack( (E, np.zeros( (E.shape[0], E.shape[0]) ) ) )
    L = np.vstack( (L1, L2) )
    b = np.concatenate( (M.T @ Rm, np.zeros((E.shape[0]))))

    Rl = linalg.solve(L, b)
    R = Rl[0:E.shape[1]]

    return R
```

```
for iD in range( 8 ):    # select datasets
    print( 'D = {:5.2}'.format( D[iD] ) )
    print( 'measured rates   :\t', end='' )
    print( (';{:6.4f}'*6).format( *Rm[:,iD ] ) )
    iRr = lagrange_solve( E, M, Rm[:,iD] )
    Rr[:,iD] = iRr
    print( 'reconciled rates :\t', end='' )
    print( (';{:06.4f}'*6).format( *Rr[:,iD ] ) )
    print( 'change           :\t', end='' )
    print( (';{:6.2f}'*6).format( *(Rr[:,iD ]/Rm[:,iD] * 100 - 100) ) )
    print( '')
```

	Substrate	Oxygen	Nitrogen	Biomass	Ethanol	CO2
D = 0.05						
measured rates :	-0.0075	-0.0167	-0.0054	0.0269	0	0.0181
reconciled rates :	-0.0075	-0.0167	-0.0054	0.0269	0	0.0181
change :	-0.04	-0.01	0.01	0		0.01
D = 0.1						
measured rates :	-0.015	-0.0335	-0.0108	0.0538	0	0.0362
reconciled rates :	-0.015	-0.0335	-0.0108	0.0538	0	0.0362
change :	-0.03	0.01	0.02	0		-0.01
D = 0.15						
measured rates :	-0.0225	-0.0502	-0.0161	0.0807	0	0.0543
reconciled rates :	-0.0225	-0.0502	-0.0161	0.0807	0	0.0542
change :	-0.01	0.02	0.04	0		-0.02
D = 0.25						
measured rates :	-0.0375	-0.0837	-0.0269	0.1343	0	0.0904
reconciled rates :	-0.0375	-0.0837	-0.0269	0.1344	0	0.0904
change :	-0.04	0.01	-0.01	0		-0.01
D = 0.28						
measured rates :	-0.042	-0.0792	-0.0259	0.1296	0.0079	0.0979
reconciled rates :	-0.041	-0.0806	-0.026	0.1299	0.0096	0.0967
change :	-2.41	1.71	0.17	0.18	21.4	-1.21
D = 0.3						
measured rates :	-0.0449	-0.0529	-0.0186	0.0928	0.0266	0.0973
reconciled rates :	-0.0418	-0.0571	-0.0187	0.0935	0.0318	0.0936
change :	-6.98	7.93	0.81	0.76	19.72	-3.78
D = 0.33						
measured rates :	-0.0493	-0.0371	-0.0143	0.0716	0.0422	0.1019
reconciled rates :	-0.0448	-0.0432	-0.0145	0.0727	0.0498	0.0966
change :	-9.18	16.34	1.43	1.42	17.95	-5.21
D = 0.35						
measured rates :	-0.0522	-0.0317	-0.013	0.0649	0.05	0.1061
reconciled rates :	-0.0472	-0.0385	-0.0132	0.066	0.0584	0.1002
change :	-9.66	21.36	1.84	1.74	16.91	-5.59

Table 3: Comparison of reconciled and measured (original) rates. For low dilution rates (D <=0.25) there are hardly corrections. For D > 0.25, Substrate uptake and ethanol are the most affected rates.

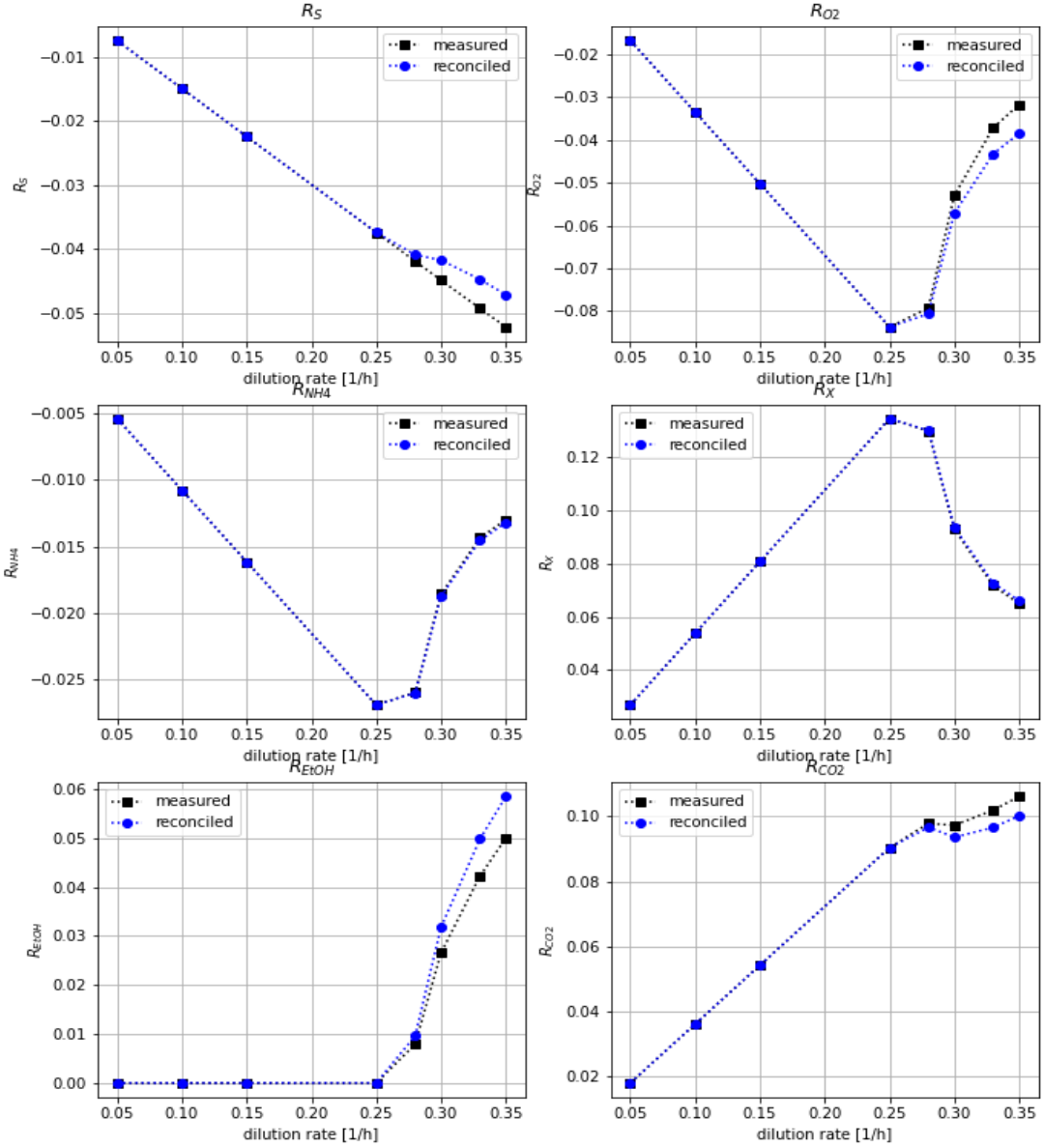

Figure 11: Measured (black) vs reconciled (blue) rates for the different dilution rates.

Reconciled and measured rates agree well for dilution rates ≤ 0.25 h^{-1}. For higher dilution rates, several rates deviate, indicating that there is a systematic deviation and the assumed process model rates are insufficient to describe the process, there is a gap in the balances that cannot only be explained by measurement errors, but it is systematic.

3.4 Hidden processes

The previous reconciliation has shown that there is redundancy – this can be used to increase the accuracy of the rate calculation, or, to estimate a rate that is not measured.

Let's assume that in the described process, ethanol evaporation takes place. Thus, next to the 6 rates taken into account, there is now another rate, $R_{EtOH,ev}$ describing the amount of ethanol leaving the reactor (per hour). This rate is now added to the vector R – while there are no new measurements, nor new experimental variables ($\mathbf{R}_m$ and $\mathbf{R}_e$ stay the same). The matrix $\mathbf{M}_m$, and $\mathbf{E}$ have to adjust to the extended vector (i.e. add a column for $R_{EtOH,ev}$). The evaporation has no impact on the liquid ethanol outflow, nor another variable, $\mathbf{M}_m$ receives a 0 column at the end. The evaporation of ethanol does have an impact on the carbon and electron balance, i.e. the matrix $\mathbf{E}$ will be extended with $(2, 12, 0)^T$ for the C, e- and N content.

$$\mathbf{R} = \begin{pmatrix} -R_S \\ -R_{O2} \\ -R_{NH4} \\ R_X \\ R_{EtOH} \\ R_{CO2} \\ R_{EtOH,ev} \end{pmatrix} \quad \mathbf{R}_m = \begin{pmatrix} -R_{m,S} \\ -R_{m,O2} \\ -R_{m,NH} \\ R_{m,X} \\ R_{m,EtOH} \\ R_{m,CO2} \end{pmatrix} \mathbf{M}_m = \begin{pmatrix} 1 & & & & & & 0 \\ & 1 & & & & & 0 \\ & & 1 & & & & 0 \\ & & & 1 & & & 0 \\ & & & & 1 & & 0 \\ & & & & & 1 & 0 \end{pmatrix} \tag{54}$$

Note that the uptake rates are made negative in the $\mathbf{R}$ vector, this way the $\mathbf{E}$ matrix can have positive entries and the balances take up- and excretion rates correctly into account (assuming that the rate was defined positive, otherwise the minus is not needed, i.e. Rm are also defined negative for uptake).

$$\mathbf{E} = \begin{pmatrix} 6 & 0 & 0 & 1 & 2 & 1 & 2 \\ 24 & -4 & 0 & 4.2 & 12 & 0 & 12 \\ 0 & 0 & 1 & 0.2 & 0 & 0 & 0 \end{pmatrix} \tag{55}$$

This can now be applied to the whole dataset:

```
# get element composition for extended rates, i.e. including rate for evaporated etoh (from excel sheet)
df = pd.read_excel(path, header=[0], sheet_name="C_N_e_ev", index_col=[0] ) Emat = np.array( df )
# measurement vector:
Rm = np.vstack( (-Rin, Rout) )
# select columns as sorted in R:
E = Emat[ :,[2,7,3,0,1,6,8] ]
M = np.eye( 6 )
Mev = np.hstack( (M, np.zeros( (M.shape[1],1) ) ) )
Rr = np.zeros( (E.shape[1], Rm.shape[1]) ) # columns of E -> # of rates
for iD in range( 8 ): # select datasets
    print( 'D = {:5.2}'.format( D[iD] ) )
    print( 'measured rates \t', end='' )
    print( (';{:6.4f}'*6).format( *Rm[:,iD ] ) )
    iRr = lagrange_solve( E, Mev, Rm[:,iD] )
    Rr[:,iD] = iRr
    print( 'reconciled rates \t', end='' )
    print( (';{:06.4f}'*7).format( *Rr[:,iD ] ) )
    print( 'change \t', end='' )
    print( (';{:6.2f}'*6).format( *(Rr[:-1,iD ]/Rm[:,iD] * 100 - 100) ) )  print( '')
```

	Substrate	Oxygen	Nitrogen	Biomass	Ethanol	CO2	EtOH evaporation
D = 0.28							
measured rates	-0.042	-0.079	-0.026	0.13	0.008	0.098	
reconciled rates	-0.042	-0.079	-0.026	0.13	0.008	0.098	0.004
change	0	0	0	0	0	0	
D = 0.3							
measured rates	-0.045	-0.053	-0.019	0.093	0.027	0.097	
reconciled rates	-0.045	-0.053	-0.019	0.093	0.027	0.097	0.013
change	0	0	0.1	0	0	0	
D = 0.33							
measured rates	-0.049	-0.037	-0.014	0.072	0.042	0.102	
reconciled rates	-0.049	-0.037	-0.014	0.072	0.042	0.102	0.019
change	0	0	0	0	0	0	
D = 0.35							
measured rates	-0.052	-0.032	-0.013	0.065	0.05	0.106	
reconciled rates	-0.052	-0.032	-0.013	0.065	0.05	0.106	0.021
change	0	0	0	0	0	0	

Table 4: Reconciled and measured rates for D > 0.25, including the unmeasured process of ethanol evaporation.

With the implementation of ethanol evaporation, the reconciled and measured rates now also agree well for dilution rates > 0.25 h^{-1}.

Additionally, there is an expected linear relation for the estimated evaporation rate and the broth ethanol concentration.

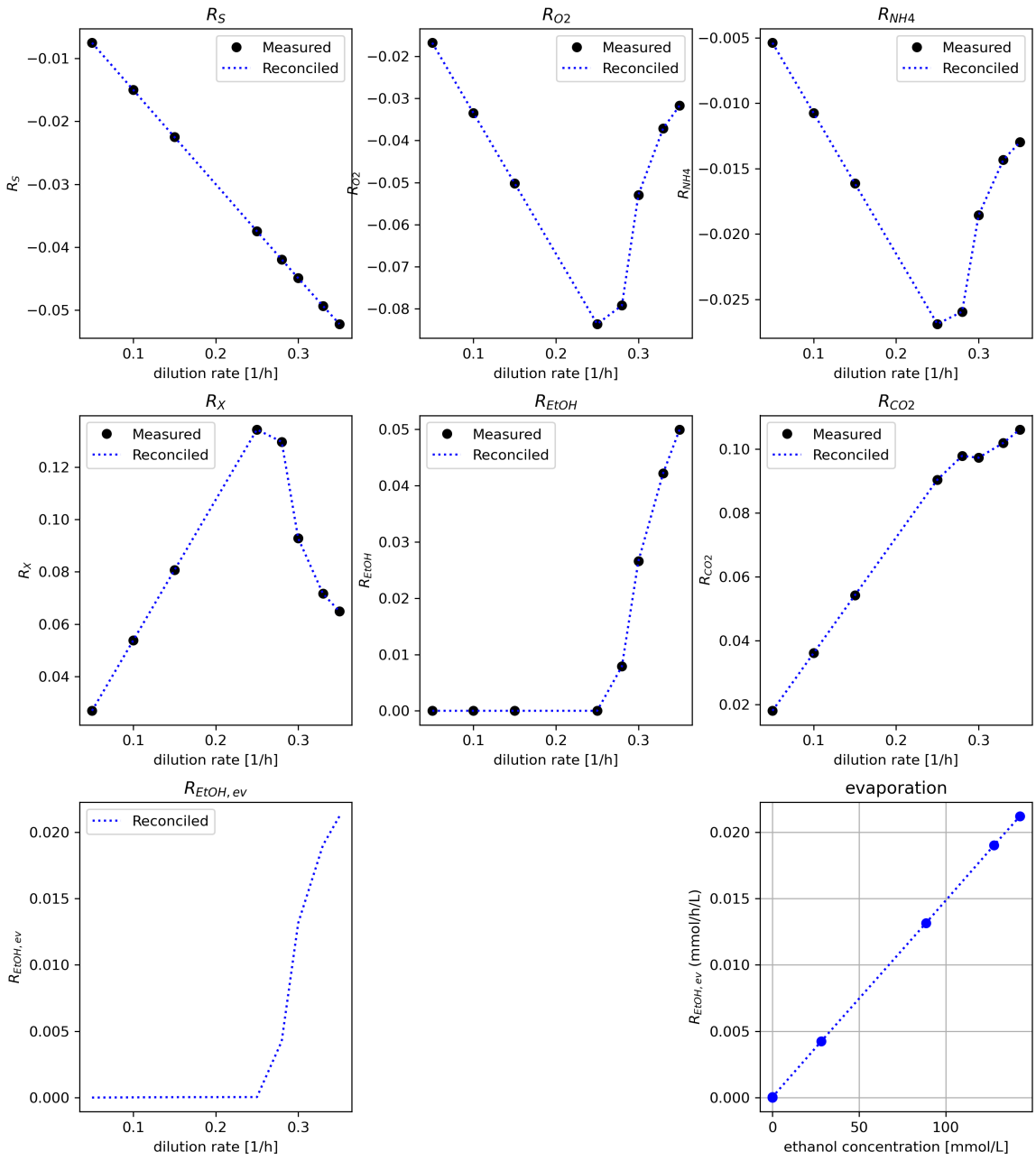

Figure 12: Reconciled and measured rates (plots 1-7) including ethanol evaporation as hidden process. Plot 9 : Estimated ethanol evaporation rate as a function of the residual ethanol concentration.

3.5 Further improvement by including TOC

The total organic carbon (TOC) measurements contain information about the carbon content in the broth, with either including biomass (TOC broth), or not (TOC filtrate). These measurements are beneficial to identify if there are unmeasured compounds in the broth, like an unknown product, or lysed biomass (complex mixture of proteins, lipids, etc).

TOC is an additional measurement, not an element balance, or additional process rate. Thus, this information should enter the measurement matrix, and the TOC has to be coupled to the process rates (the variables **R**).

Therefore we need to derive the relation. TOC is a carbon concentration measurement that can be expressed as a function of the compounds in the broth (here biomass, residual glucose, ethanol):

$$TOC_b = 1c_X + 2c_{EtOH} + 6c_{Glc} \tag{56}$$

The concentration of biomass, ethanol and residual glucose can be derived from the process rates using the bioreactor balances:

$$\begin{aligned}
&\frac{dc_X V_L}{dt} = R_X V_L - c_X F_{\text{out}} \qquad c_X = \frac{R_X V_L}{F_{\text{out}}} = \frac{R_X}{D} \\
&\frac{dc_{EtOH} V_L}{dt} = R_{EtOH} V_L - c_{EtOH} F_{\text{out}} \qquad c_{EtOH} = \frac{R_{EtOH} V_L}{F_{\text{out}}} = \frac{R_{EtOH}}{D} \\
&\frac{dc_S V_L}{dt} = c_{S,\,\text{in}}\, F_{\text{in}} - R_S V_L - c_S F_{\text{out}} \qquad c_S = \frac{c_{S,\,\text{in}}\, F_{\text{in}} - R_S V_L}{F_{\text{out}}} \quad \text{with} \quad F_{\text{in}} = F_{\text{out}} : c_S = \frac{c_{S,\,\text{in}}\, D - R_S}{D}
\end{aligned} \tag{57}$$

The TOC equation can be rewritten as:

$$\begin{aligned}
TOC_b &= 1c_X + 2c_{EtOH} + 6c_{Glc} \\
TOC_b &= 1\frac{R_X}{D} + 2\frac{R_{EtOH}}{D} + 6\frac{Dc_{S,in} - R_S}{D} \\
TOC_b - 6c_{S,\,\text{in}} &= 1\frac{R_X}{D} + 2\frac{R_{EtOH}}{D} - 6\frac{R_S}{D}
\end{aligned} \tag{58}$$

Similarly, TOC filtrate can be derived, without all the details, the measurement reads:

$$\begin{aligned}
TOC_f &= 2c_{EtOH} + 6c_{Glc} \\
TOC_f - 6c_{S,in} &= 2\frac{R_{EtOH}}{D} + 6\frac{R_S}{D}
\end{aligned} \tag{59}$$

This should now be integrated in the measurement equation system. As in all other systems, known and unknown are separated, (i.e. cS, in has to move to the LHS):

$$\begin{aligned}
D\,(TOC_b - 6c_{S,\,\text{in}}) &= R_X + 2R_{EtOH} - 6R_S \\
D\,(TOC_f - 6c_{S,\,\text{in}}) &= 2R_{EtOH} - 6R_S
\end{aligned} \tag{60}$$

Leading to a new measurement matrix **M**:

$$\mathbf{R} = \begin{pmatrix} -R_S \\ -R_{O2} \\ -R_{NH4} \\ R_X \\ R_{EtOH} \\ R_{CO2} \\ R_{EtOH,ev} \end{pmatrix} \quad \mathbf{R}_m = \begin{pmatrix} -R_{m,S} \\ -R_{m,O2} \\ -R_{m,NH} \\ R_{m,X} \\ R_{m,EtOH} \\ R_{m,CO2} \\ D\,(TOC_b - 6c_{S,\,\text{in}}\,) \\ D\,(TOC_f - 6c_{S,\,\text{in}}\,) \end{pmatrix} \quad \mathbf{M}_m = \begin{pmatrix} 1 & & & & & & 0 \\ & 1 & & & & & 0 \\ & & 1 & & & & 0 \\ & & & 1 & & & 0 \\ & & & & 1 & & 0 \\ & & & & & 1 & 0 \\ 6 & & & 1 & 2 & & \\ 6 & & & & 2 & & \end{pmatrix} \tag{61}$$

(62)

NOTE: The sign for R_S in M changes as in the **R** vector definition, R_S is negative.

```
Rm = np.vstack( (-Rin, Rout) )

# select columns as sorted in R:
E = Emat[ :,[2,7,3,0,1,6,8] ]

M = np.eye( 6 )
Mev = np.hstack( (M, np.zeros( (M.shape[1],1) )  ) )
MevTOC = np.vstack( (Mev, [6, 0, 0, 1, 2, 0, 0], [6, 0, 0, 0, 2, 0, 0]  ) )

cSin = 150e-3 #mmol/L

# add TOC_b and TOC_f measurements to Rm
RmTOC = np.vstack( (Rm, D*data[:,5]*1e-3 - 6*D*cSin, D*data[:,6]*1e-3 - 6*D*cSin) )

Rr = np.zeros( (E.shape[1], Rm.shape[1]) )       # columns of E -> # of rates

for iD in range( Rm.shape[1] ):    # select datasets
    print( 'D = {:5.2}'.format( D[iD] ) )
    print( 'measured rates   \t', end='' )
    print( (';{:6.4f}'*8).format( *RmTOC[:,iD ] ) )
    iRr = lagrange_solve( E, MevTOC, RmTOC[:,iD] )
    Rr[:,iD] = iRr
    print( 'reconciled rates \t', end='' )
    print( (';{:06.4f}'*7).format( *Rr[:,iD ] ) )
    print( 'change           \t', end='' )
    print( (';{:6.2f}'*6).format( *(Rr[:-1,iD ]/Rm[:,iD] * 100 - 100) ) )
    print( '')
```

Here the results are very comparable to the earlier calculations. The balances were already well closed, the TOC broth and filtrate confirm the earlier calculation and hypothesis of a compound in the gas-phase (i.e. ethanol evaporation).

Additional to the reconciliation itself, the accuracy of the estimated rates can be obtained from error propagation.

With the use of Lagrange multipliers this is not straight forward (though logical), and is therefore included in a python script that is provided on BrightSpace. Only additional input to the function are the respective measurement standard deviations (std_dev_y).

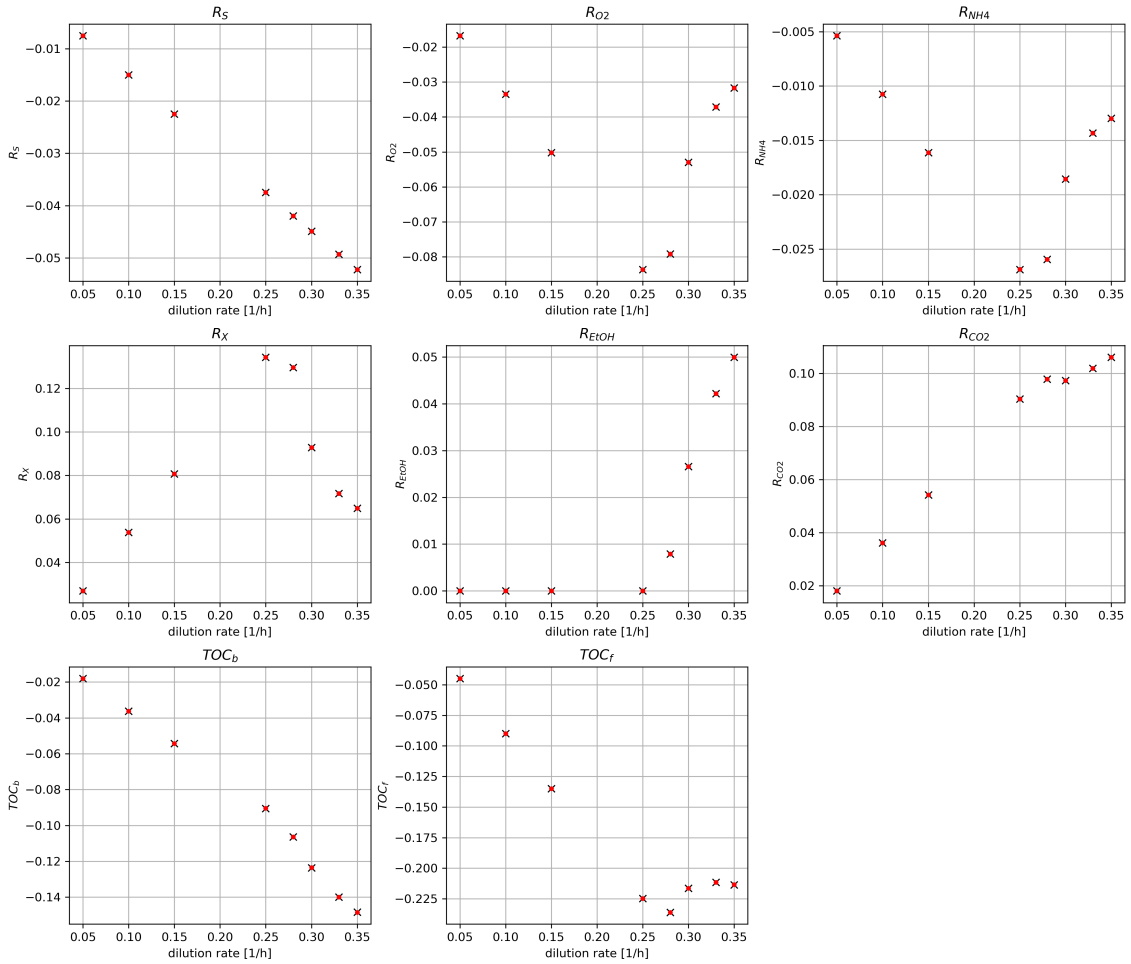

Figure 13: Reconciled and measured rates and additional observations from TOC_f and TOC_b. Red : reconciled, blue cross : measurement.

3.6 Error propagation

Biological data is typically noisy data, and for decision making it is relevant to discriminate between accurate and less accurate rates, parameters etc. For our calculations, two aspects will be taken into account:

- The accuracy of each point should be taken into account during the regression (weighted regression). I.e. a known, well determined measurement should have more “weight" in the calculation than highly noisy data points.
- The accuracy of the obtained parameters – which are reliable, which less, and is there correlation (we will not look too much into that though).

This follows a "standard" rescaling approach (i.e. every measurement equation is divided by the respective standard

deviation). A Lagrange optimization with weighting is defined and can be used for the calculations (assignments, exam, etc).

This code takes as inputs the matrices and vectors as before, adding a vector containing the standard deviations of the observations Rm.

```
def lagrange_solve_w( E, M, Rm, std_dev_Rm ):
    # inputs:
    # matrix E (n x m) - linearity constraints (m balance equations E . R = 0)
    # matrix M (k x m) - measurement matrix for the k measurements (M . R = Rm +/-)
    # vector Rm (m) - measurement values
    # vector std_dev_Rm - standard deviations of the measurement in Rm
    # rescale equation according to std_dev_Rm
    M_w = np.diag( std_dev_Rm ** -1 ) @ M
    L1 = np.hstack( (M_w.T @ M_w, E.T) )
    L2 = np.hstack( (E, np.zeros( (E.shape[0], E.shape[0]) ) ) )
    L = np.vstack( (L1, L2) )
    b = np.concatenate( (M_w.T @ (Rm/std_dev_Rm), np.zeros((E.shape[0]))))
    Rl = linalg.solve(L, b)
    R = Rl[0:E.shape[1]]

    # error propagation
    M = np.pad( M_w, ((0,0),(0,b.shape[0]-M_w.shape[1])), mode='constant')
    J = linalg.inv(L) @ M.T
    S_R = J @ J.T
    std_dev_R = np.diag(S_R) ** 0.5
    std_dev_R = std_dev_R[0:E.shape[1]]
    return (R, std_dev_R)

# add TOC_b and TOC_f measurements to Rm
# assume 5% error on the rate observations (simplified)
std_dev_RmTOC = 0.05 * np.abs( RmTOC )
# make sure, that it is never too close to 0
std_dev_RmTOC[ np.where(std_dev_RmTOC<1e-6)] = 1e-6

Rr = np.zeros( (E.shape[1], Rm.shape[1]) ) # columns of E ->
# of rates
sd_Rr = np.zeros( (E.shape[1], Rm.shape[1]) ) # columns of E -> # of rates

for iD in range( Rm.shape[1] ): # select datasets
    print( 'D = {:5.2}'.format( D[iD] ) )
    print( 'measured rates ', end='' )
    print( ('\t{:6.4f}'*8).format( *RmTOC[:,iD ]*1e3 ) )
    (iRr, std_dev_iRr) = lagrange_solve_w( E, MevTOC, RmTOC[:,iD], std_dev_RmTOC[:,iD] )  Rr[:,iD] = iRr
    sd_Rr[:,iD] = std_dev_iRr
    print( 'reconciled rates ', end='' )
    print( ('\t{:6.4f} '*7 ).format( *Rr[:,iD ]*1e3 ) )
    print( 'standard deviation ', end='' )
    print( ('\t{:6.4f} '*7 ).format( *std_dev_iRr*1e3 ) )
    print( '')

# generate a figure showing the results (measured vs reconciled as a function of dilution rate)
Rmtxt = [ '$R_S$', '$R_{O2}$', '$R_{NH4}$', '$R_{X}$', '$R_{EtOH}$', '$R_{CO2}$', '$TOC_b$', '$TOC_f$']

# generate data plots
plt.figure( figsize = (16, 14 ))

for i in range( RmTOC.shape[0] ):
    plt.subplot(3, 3, i+1)
    #plt.plot( D, RmTOC[i,:]*1e3, 'kx' )
    plt.errorbar(D, RmTOC[i,:]*1e3, std_dev_RmTOC[i,:]*1e3, marker='x', mec='black', ms=8, mew=1, ls = '')
    for n in range( RmTOC.shape[1]): # loop over all experiments, calculate expected measurement
        Rme = MevTOC @ Rr[:,n]
        sd_Rme = MevTOC @ sd_Rr[:,n]
        plt.errorbar(D[n], Rme[i]*1e3, sd_Rme[i]*1e3, marker='.', mfc='red', ms=8, mew=0, ls ='')
        #plt.plot( D[n], Rme[i]*1e3, 'r.' )

```

```
62      plt.xlabel( 'dilution rate [1/h]' )
63      plt.ylabel( 'mmol/L/h' )
64      plt.title( Rmtxt[i] )
65      plt.grid()
66
67  plt.subplots_adjust( wspace = 0.25, hspace = 0.25 )
68  plt.savefig( 'Sc_process_rates_recon_TOC.png', dpi=300 )
69  plt.show()
```

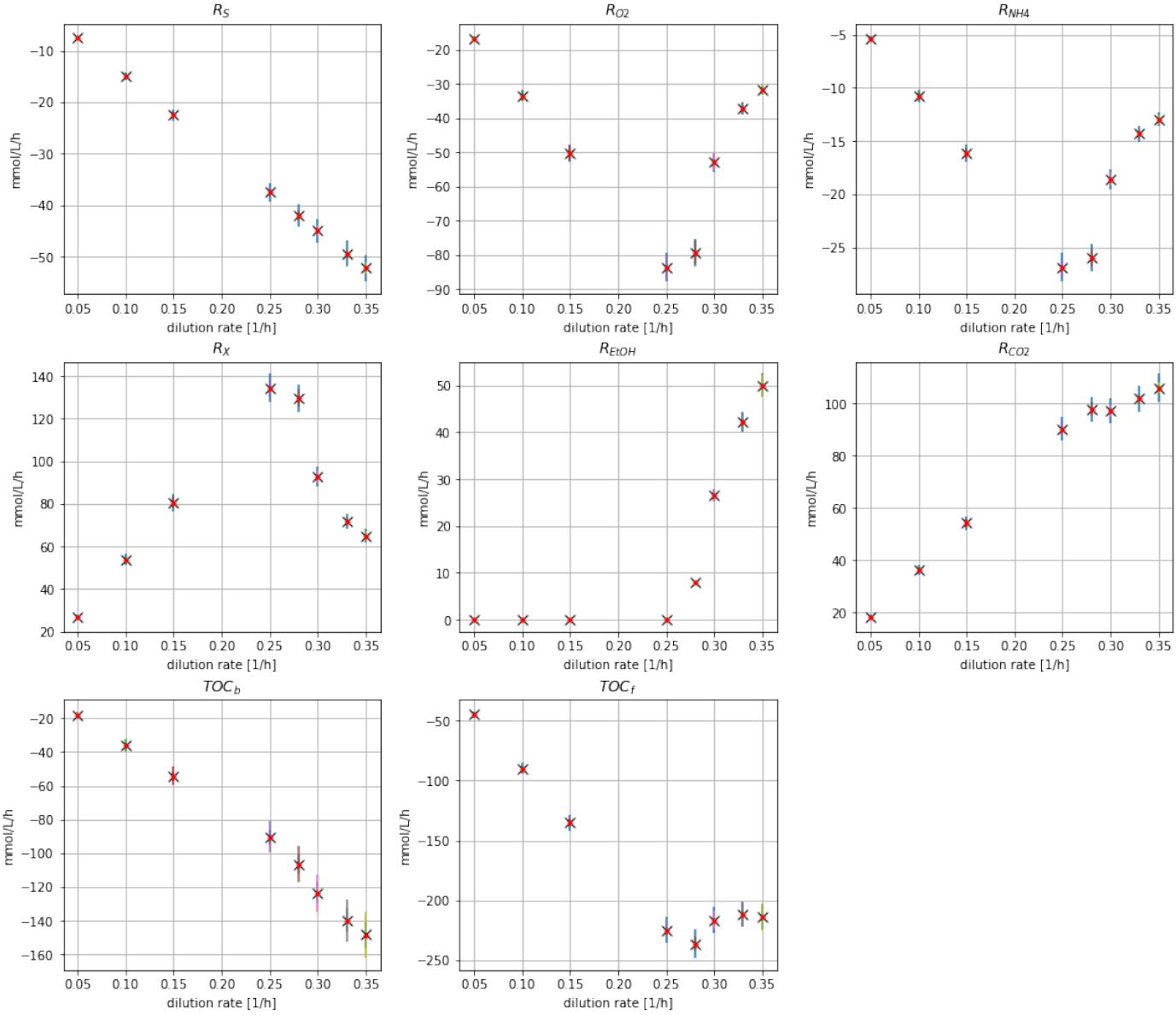

Figure 14: reconciled rates with respective standard deviation from data reconciliation.