Escherichia coli Core Metabolism Model in LIM

Karline Soetaert
Centre for Estuarine and Marine Ecology
Netherlands Institute of Ecology
The Netherlands

Abstract

R package **LIM** (Soetaert and van Oevelen 2009a) is designed for reading and solving linear inverse models (LIM).

A package vignette deals with the structure of the LIM input files and how to solve the problems (Soetaert and van Oevelen 2009b).

To open it, type:

`vignette("LIM")`

Here it is exemplified on a (relatively small) problem from systems biology, the core metabolism of E. coli (Edwards, Covert, and Palsson 2002) as from the following website:

`http://gcrg.ucsd.edu/Downloads/Flux_Balance_Analysis`

The original input file can be found in the package subdirectory `/examples/Reactions/E_coli.lim`

If you use this package, please cite as: (van Oevelen, van den Meersche, Meysman, Soetaert, Middelburg, and Vezina 2009).

Keywords: Linear inverse models, flux balance analysis, linear programming, E coli, R.

1. the E. coli input file

The input file consists of several sections (see package vignette).

- The header of the file (ends at first line with ###) is ignored
- The metabolic reactions
- A function to maximise
- The bounds on the reactions (inequalities)
- A measurement equation
- The name of the components
- The names of the externals

Everything following a "!" is ignored.

```
E.coli input file
###############################################


## REACTIONS
!gene:  Reaction                                        enzyme
GLK1:  GLC + ATP -> G6P + ADP                    !      Glucokinase
PGI1:  G6P <-> F6P                               !      Phosphoglucose isomerase
PFKA:  F6P + ATP -> FDP + ADP                    !      Phosphofructokinase
FBP:   FDP -> F6P + PI                           !      Fructose-1,6-bisphosphatase
FBA:   FDP <-> T3P1 + T3P2                       !      Fructose-1,6-bisphosphatate al
TPIA:  T3P2 <-> T3P1                             !      Triosphosphate Isomerase
GAPA:  T3P1 + PI + NAD <->  NADH + 13PDG         !      Glyceraldehyde-3-phosphate deh
PGK:   13PDG + ADP <-> 3PG + ATP                 !      Phosphoglycerate kinase
GPMA:  3PG <-> 2PG                               !      Phosphoglycerate mutase 1
ENO:   2PG <-> PEP                               !      Enolase
PPSA:  PYR + ATP -> PEP + AMP + PI               !      Phosphoenolpyruvate synthase
PYKA:  PEP + ADP -> PYR + ATP                    !      Pyruvate Kinase II
!      PYKF: PEP + ADP -> PYR + ATP              !      Pyruvate Kinase I
ACEE:  PYR + COA + NAD -> NADH + CO2 + ACCOA     !      Pyruvate dehydrogenase

!Pentose Phosphate Pathway
ZWF:   G6P + NADP <-> D6PGL + NADPH              !      Glucose 6-phosphate-1-dehydrog
PGL:   D6PGL -> D6PGC                            !      6-Phosphogluconolactonase
GND:   D6PGC + NADP -> NADPH + CO2 + RL5P        !      6-Phosphogluconate dehydrogena
RPIA:  RL5P <-> R5P                              !      Ribose-5-phosphate isomerase A
RPE:   RL5P <-> X5P                              !      Ribulose phosphate 3-epimerase
TKTA1: R5P + X5P <-> T3P1 + S7P                  !      Transketolase I
!  TKTB1: R5P + X5P <-> T3P1 + S7P               !      Transketolase II
TKTA2: X5P + E4P <-> F6P + T3P1                  !      Transketolase I
!  TKTB2: X5P + E4P <-> F6P + T3P1               !      Transketolase II
TALA:  T3P1 + S7P <-> E4P + F6P                  !      Transaldolase A

!The Tricarboxylic Acid Cycle
GLTA:  ACCOA + OA -> COA + CIT                   !      Citrate synthase
ACNA:  CIT <-> ICIT                              !      Aconitase A
ICDA:  ICIT + NADP <-> CO2 + NADPH + AKG         !      Isocitrate dehydrogenase
SUCA:  AKG + NAD + COA -> CO2 + NADH + SUCCOA    !      2-Ketoglutarate dehyrogenase
SUCC1: SUCCOA + ADP + PI <-> ATP + COA + SUCC    !      Succinyl-CoA synthetase
SDHA1: SUCC + FAD -> FADH + FUM                  !      Succinate dehydrogenase
FRDA:  FUM + FADH -> SUCC + FAD                  !      Fumurate reductase
```

```
FUMA:   FUM <-> MAL                                   !      Fumarase A
MDH:    MAL + NAD <-> NADH + OA                       !      Malate dehydrogenase

!Pyruvate Metabolism
DLD1:   PYR + NADH <-> NAD + LAC                      !      D-Lactate dehydrogenase 1
ADHE2:  ACCOA +2*NADH <-> ETH +2*NAD + COA            !      Acetaldehyde dehydrogenase
PFLA:   PYR + COA -> ACCOA + FOR                      !      Pyruvate formate lyase 1
PTA:    ACCOA + PI <-> ACTP + COA                     !      Phosphotransacetylase
ACKA:   ACTP + ADP <-> ATP + AC                       !      Acetate kinase A
ACS:    ATP + AC + COA -> AMP + PPI + ACCOA           !      Acetyl-CoA synthetase

!Anaplerotic Reactions
PCKA:   OA + ATP -> PEP + CO2 + ADP                   !      Phosphoenolpyruvate carboxykinase
PPC:    PEP + CO2 -> OA + PI                          !      Phosphoenolpyruvate carboxylase
MAEB:   MAL + NADP -> CO2 + NADPH + PYR               !      Malic enzyme (NADP)
SFCA:   MAL + NAD -> CO2 + NADH + PYR                 !      Malic enzyme (NAD)
ACEA:   ICIT -> GLX + SUCC                            !      Isocitrate lyase
ACEB:   ACCOA + GLX -> COA + MAL                      !      Malate synthase A
PPA:    PPI -> 2*PI                                   !      Inorganic pyrophosphatase
GLPK:   GL + ATP -> GL3P + ADP                        !      Glycerol kinase
GPSA1:  GL3P + NADP <-> T3P2 + NADPH                  !      Glycerol-3-phosphate-dehydrogenase-
RBSK:   RIB + ATP -> R5P + ADP                        !      Ribokinase

!Respiration                Note: the P/O ratio is set to 1.5 currently
NUOA:   NADH + Q -> NAD + QH2 +3.5*HEXT               !      NADH dehydrogenase I
FDOH:   FOR + Q -> QH2 + CO2 +2*HEXT                  !      Formate dehydrogenase-O
GLPD:   GL3P + Q -> T3P2 + QH2                        !      Glycerol-3-phosphate dehydrogenase
CYOA:   QH2 +0.5*O2 -> Q + 2.5*HEXT                   !      Cytochrome oxidase bo3
SDHA2:  FADH + Q <-> FAD + QH2                        !      Succinate dehydrogenase complex
PNT1A:  NADPH + NAD -> NADP + NADH                    !      Pyridine nucleotide transhydrogenase
PNT2A:  NADP + NADH +2*HEXT -> NADPH + NAD            !      Pyridine nucleotide transhydrogenase
ATPA:   ATP <-> ADP + PI +4*HEXT                      !      F0F1-ATPase

!Membrane Transport
GLCUP:  GLCxt + HEXT -> GLC                           !      Glucose/galactose transporter
GLCPTS:GLCxt + PEP -> G6P + PYR                       !      Glucose
GLUP:   GLxt <-> GL                                   !      Glycerol
RIBUP:  RIBxt + ATP <-> RIB + ADP + PI                 !      Ribose
ACUP:   ACxt + HEXT <-> AC                            !      Acetate transport
LACUP:  LACxt + HEXT <-> LAC                          !      L-Lactate
FORUP:  FORxt <-> FOR                                 !      Formate transport
ETHUP:  ETHxt + HEXT <-> ETH                          !      Ethanol transport
SUCCUP:SUCCxt + HEXT <-> SUCC                         !      Succinate transport
```

```
PYRUP: PYRxt + HEXT <-> PYR                     !     Pyruvate transport
PIUP:  PIxt <-> PI                              !     Phosphate transport
O2TX:  O2xt <-> O2                              !     Oxygen transport
CO2TX: CO2xt <-> CO2                            !     Carbon dioxide transport

ATPM: ATP -> ADP + PI                           !     ATP drain flux for constant ma
ADK:  ATP + AMP-> 2*ADP                         !     ADK

Growth:                                                                  &
41.25*ATP +3.54*NAD +18.22*NADPH +0.2*G6P                                &
 +0.07*F6P +0.89*R5P +0.36*E4P +0.12*T3P1                                &
 +1.49*3PG +0.51*PEP +2.83*PYR +3.74*ACCOA  +1.78*OA +1.07*AKG           &
 -> 3.74*COA +41.25* ADP +41.25* PI                                      &
 +3.54* NADH +18.22* NADP +1.00* Biomass
### END REACTION

## MAXIMISE
maxgrowth: Growth
## END MAXIMISE

### INEQUALITIES
! Carbon sources...
O2TX  = [0,20]     ! Oxygen input
GLCUP = [0,10]     ! glucose input
GLUP  = [-1000,0] ! glycerol
RIBUP = [-1000,0] ! Ribose uptake   - strange!
SUCCUP= [-1000,0] ! succinate
ACUP  = [-1000,0] ! acetate
LACUP = [-1000,0] ! lactate
PYRUP = [-1000,0] ! pyruvate
! Carbon byproducts
FORup = [-1000,0] ! formate
ETHup = [-1000,0] ! ethanol
CO2TX = [-1000,0] ! CO2
! phosphate
PIUP = [-1000,1000]

SDHA1  <100
FRDA   <100
FORup+ LACUP=[-10,-10]
### END INEQUALITIES

### EQUATIONS
```

```
 ATPM = 5.87  ! Non-growth associated ATP drain flux for constant maintenance requirements
### END EQUATIONS

### COMPONENTS
GLC     ! a-D-Glucose
G6P     ! Glucose 6-phosphate
F6P     ! Fructose 6-phosphate
FDP     ! Fructose 1,6-diphosphate
T3P2    ! /DHAP Dihydroxyacetone phosphate
T3P1    ! Glyceraldehyde 3-phosphate
13PDG   ! 1,3-bis-Phosphoglycerate
3PG     ! 3-Phosphoglycerate
2PG     ! 2-Phosphoglycerate
PEP     ! Phosphoenolpyruvate
PYR     ! Pyruvate
ACCOA   ! Acetyl-CoA
CIT     ! Citrate
!  ACO    ! cis-Aconitate
ICIT    ! Isocitrate
AKG     ! a-Ketoglutarate
SUCCOA  ! Succinate CoA
SUCC    ! Succinate
FUM     ! Fumarate
MAL     ! Malate
OA      ! Oxaloacetate
!  ACAL ! Acetaldehyde
ACTP    ! Acetyl-phosphate
ETH     ! Ethanol
AC      ! Acetate
LAC     ! D-Lactate
FOR     ! Formate
D6PGL   ! D-6-Phosphate-glucono-delta-lactone
D6PGC   ! D-6-Phosphate-gluconate
RL5P    ! Ribulose 5-phosphate
X5P     ! Xylulose-5-phosphate
R5P     ! Ribose 5-phosphate
S7P     ! sedo-Heptulose
E4P     ! Erythrose 4-phosphate
RIB     ! Ribose
GLX     ! Glyoxylate
NAD     ! Nicotinamide adenine dinucleotide
NADH    ! Nicotinamide adenine dinucleotide (reduced)
NADP    ! Dihydronicotinamide adenine dinucleotide phosphate
```

```
NADPH   ! Dihydronicotinamide adenine dinucleotide phosphate (reduced)
HEXT    ! External Hydrogen Ion (Proton)
Q       ! Ubiquinone

FAD     ! Flavin adenine dinucleotide
FADH    ! Flavin adenine dinucleotide (reduced)
AMP     ! Adenosine monophosphate
ADP     ! Adenosine diphosphate
ATP     ! Adenosine triphosphate
GL3P    ! Glycerol 3-phosphate
CO2     ! Carbon dioxide
PI      ! Inorganic Phosphate
PPI     ! Pyrophosphate


O2      ! Oxygen
COA
GL
QH2     !
### END COMPONENTS

### EXTERNALS
Biomass
GLCxt
GLxt
RIBxt
ACxt
LACxt
FORxt
ETHxt
SUCCxt
PYRxt
PIxt
O2xt
CO2xt
### END EXTERNALS
```

2. Reading the E.coli input file

Assuming that the input file is called "E_coli.lim" and the working directory has been set, it can be read as follows:

```
require(LIM)
LIMEcoli <- Setup("E_coli.lim")
```

This creates a list of type `lim`, that contains all information necessary to solve the problem

3. The parsimonious and optimized solution, ranges

Once the input file has been read, we can generate the "simplest" solution, i.e. the one where $\sum(x^2)$ is minimal, where x are the unknown reaction rates. This is called the "parsimonious solution". It is common to calculate this in foodweb ecology (where it is unclear which charactersitic of a foodweb is optimized); it may be less relevant from a system's biology perspective.

Function `Ldei` estimates the parsimonious solution

```
> pars <- Ldei(LIMEcoli)
```

It makes more sense to optimize the growth. That growth has to be maximised was inputted in the file (by the ## maximize statement).

The optimal value is found by linear programming, using function `Linp`:

```
> LP <- Linp(LIMEcoli)
```

It is also simple to estimate the ranges of all unknown reaction rates:

```
> xr <- Xranges(LIMEcoli)
```

Now for every reaction rate, the "simplest", "optimal", "minimal" and "maximal" value has been estimated:

```
> data.frame(simplest = pars$X, optimal = LP$X, xr)
```

```
      simplest      optimal         min         max
GLK1 1.0000335    0.0000000   0.0000000   10.000000
PGI1 4.2838919  807.5327453 -15.8333333  807.532745
PFKA 4.4703252  781.5906859   0.8333333 2229.130000
FBP  0.1864334    0.0000000   0.0000000 1604.130000
FBA  4.2838919  781.5906859   0.8333333  781.590686
TPIA 4.2838919  781.5906859   0.8333333  781.590686
GAPA 8.5677837 1541.4341987   5.0000000 1541.434199
```

PGK	8.5677837	1541.4341987	5.0000000	1541.434199
GPMA	8.5677837	1492.0890901	5.0000000	1492.089090
ENO	8.5677837	1492.0890901	5.0000000	1492.089090
PPSA	0.3810706	0.0000000	0.0000000	1604.130000
PYKA	3.0394798	466.6579644	0.0000000	2136.630000
ACEE	0.0000000	1158.9491902	0.0000000	1158.949190
ZWF	0.0000000	0.0000000	0.0000000	75.000000
PGL	0.0000000	0.0000000	0.0000000	75.000000
GND	0.0000000	0.0000000	0.0000000	75.000000
RPIA	0.0000000	23.6238328	0.0000000	28.202015
RPE	0.0000000	-23.6238328	-23.6238328	50.000000
TKTA1	0.0000000	-5.8507623	-5.8507623	25.000000
TKTA2	0.0000000	-17.7730705	-17.7730705	25.000000
TALA	0.0000000	-5.8507623	-5.8507623	25.000000
GLTA	1.4322163	35.4357492	0.0000000	40.847149
ACNA	1.4322163	35.4357492	0.0000000	40.847149
ICDA	0.0000000	35.4357492	0.0000000	40.847149
SUCA	0.0000000	0.0000000	0.0000000	30.000000
SUCC1	0.0000000	0.0000000	0.0000000	30.000000
SDHA1	1.4322163	0.0000000	0.0000000	100.000000
FRDA	0.0000000	100.0000000	0.0000000	100.000000
FUMA	1.4322163	-100.0000000	-100.0000000	8.333333
MDH	-1.1932998	-100.0000000	-1168.3150000	16.666667
DLD1	4.2144864	9.6539066	0.0000000	10.000000
ADHE2	2.9210810	1000.0000000	0.0000000	1000.000000
PFLA	5.7855136	0.3460934	0.0000000	150.000000
PTA	0.8551376	0.0000000	0.0000000	1660.380000
ACKA	0.8551376	0.0000000	0.0000000	1660.380000
ACS	0.8551376	0.0000000	0.0000000	1604.130000
PCKA	1.0156212	0.0000000	0.0000000	1604.130000
PPC	3.6411373	194.3849394	0.0000000	1704.130000
MAEB	1.1959311	0.0000000	0.0000000	1068.315000
SFCA	2.8618012	0.0000000	0.0000000	1068.315000
ACEA	1.4322163	0.0000000	0.0000000	30.000000
ACEB	1.4322163	0.0000000	0.0000000	30.000000
PPA	0.8551376	0.0000000	0.0000000	1604.130000
GLPK	0.0000000	0.0000000	0.0000000	0.000000
GPSA1	-1.3755677	0.0000000	-140.0000000	0.000000
RBSK	0.0000000	0.0000000	0.0000000	0.000000
NUOA	0.0000000	140.0000000	0.0000000	140.000000
FDOH	0.0000000	0.0000000	0.0000000	140.000000
GLPD	1.3755677	0.0000000	0.0000000	140.000000
CYOA	2.8077840	40.0000000	0.0000000	40.000000

```
SDHA2    1.4322163  -100.0000000  -100.0000000    8.333333
PNT1A    1.6658701     0.0000000     0.0000000 3208.260000
PNT2A    1.8455067   567.9655123     0.0000000 3208.260000
ATPA    -2.3659951  -145.4663293  -460.0000000 1144.130000
GLCUP    1.0000335     0.0000000     0.0000000   10.000000
GLCPTS   3.2838584   814.1562498     0.0000000  814.156250
GLUP     0.0000000     0.0000000     0.0000000    0.000000
RIBUP    0.0000000     0.0000000     0.0000000    0.000000
ACUP     0.0000000     0.0000000   -75.0000000    0.000000
LACUP   -4.2144864    -9.6539066   -10.0000000    0.000000
FORUP   -5.7855136    -0.3460934   -10.0000000    0.000000
ETHUP   -2.9210810 -1000.0000000 -1000.0000000    0.000000
SUCCUP   0.0000000  -100.0000000  -130.0000000    0.000000
PYRUP    0.0000000   -18.1424351  -150.0000000    0.000000
PIUP     0.0000000   120.5477822     0.0000000  120.547782
O2TX     1.4038920    20.0000000     0.0000000   20.000000
CO2TX   -1.4322163 -1000.0000000 -1000.0000000    0.000000
ATPM     5.8700000     5.8700000     5.8700000    5.870000
ADK      1.2362082     0.0000000     0.0000000 1604.130000
Growth   0.0000000    33.1175226     0.0000000   33.117523
```

The range solutions can be plotted; as there are many reactions, we plot them in two figures. The "optimal" solution is added as a black dot.

```
> par(mfrow = c(1, 2))
> nr <- LIMEcoli$NUnknowns
> ii <- 1:(nr/2)
> dotchart(LP$X[ii], xlim = range(xr), pch = 16, cex = 0.8)
> segments(xr[ii,1], 1:nr, xr[ii,2], 1:nr)
> ii <- (nr/2+1):nr
> dotchart(LP$X[ii], xlim = range(xr), pch = 16, cex = 0.8)
> segments(xr[ii,1], 1:nr, xr[ii,2], 1:nr)
> mtext(side =  3, cex = 1.5, outer = TRUE, line = -1.5,
+       "E coli Core Metabolism, optimal solution and ranges")
```

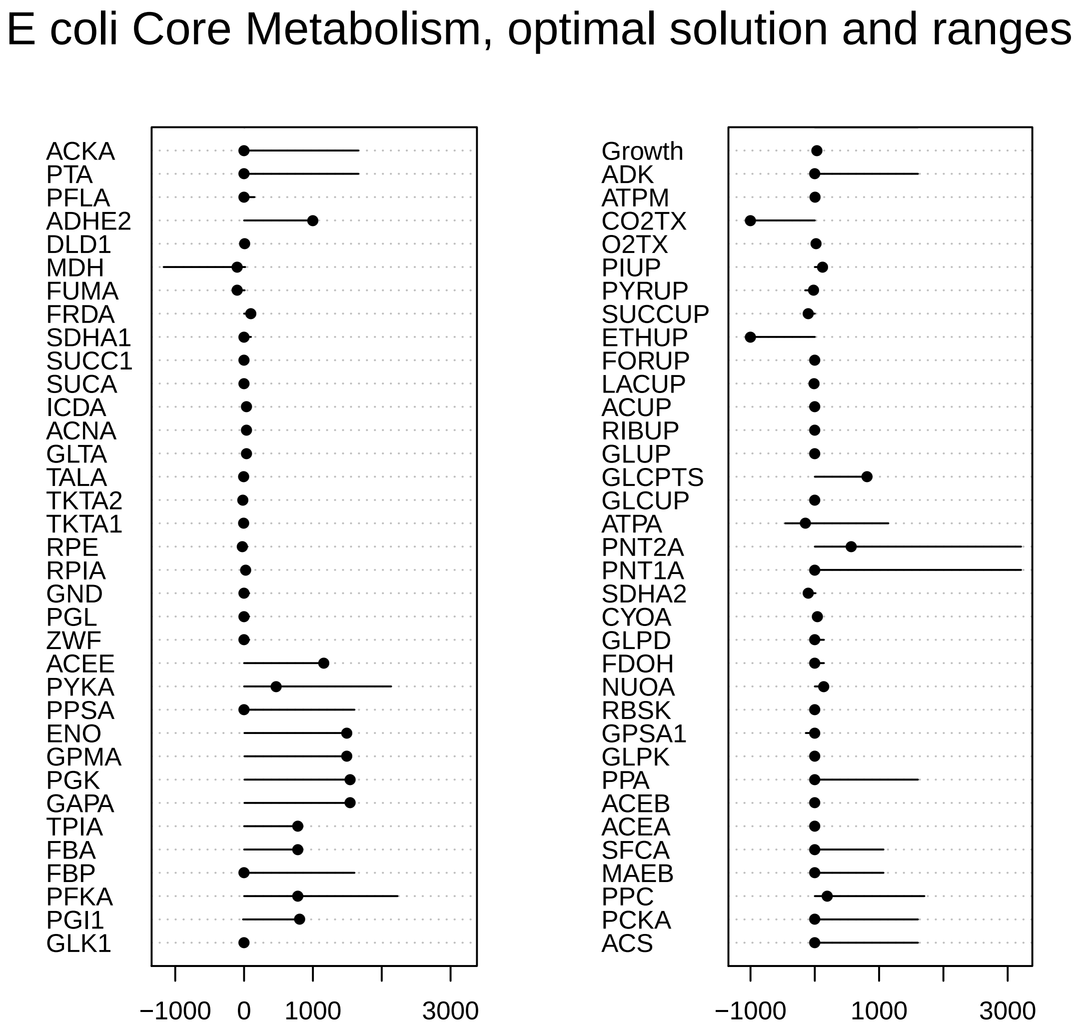

Figure 1: Ranges, and optimal solution of the E.coli central core metabolism - see text for R-code

4. Generating multiple plausible solutions

The E.coli model is underdetermined, such that an infinite amount of solutions are likely, given the data. By optimising growth, we selected one "optimal" solution; by estimating the ranges, we calculated the minimal and maximal values of each reaction.

It is also possible to sample the solution space in a random way. Function `xsample` does that; each point it generates is equally valid and equally likely.

We take 500 random samples; it takes a while to do this; `print(system.time()` estimates the time, in seconds.

```
> print(system.time(
+   xs <- Xsample(LIMEcoli, iter = 500, type = "mirror", test = TRUE)  #))
+ ))
```

```
   user  system elapsed 
   4.75    0.30    5.06 
```

With 70 variables, it is not possible to plot all pairwise relationships.

Here we plot them for 12 of them.

```
>   pairs(xs[, 1:12], pch = ".", cex = 2, gap = 0, upper.panel = NULL)
```

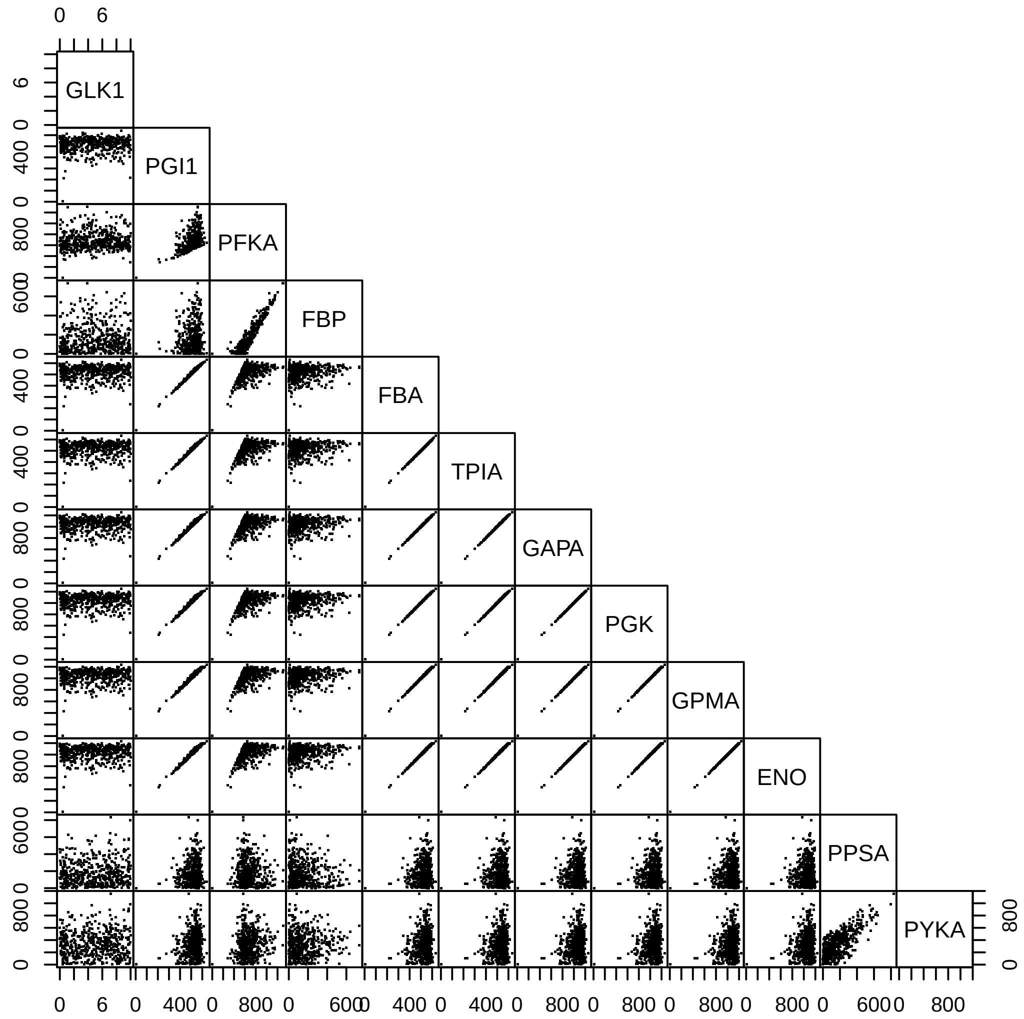

Figure 2: A random sample of plausible solutions of the E.coli central core metabolism - plotted as a pairwise plot for the first 12 reaction rates see text for R-code

References

Edwards J, Covert M, Palsson B (2002). "Metabolic Modeling of Microbes: the Flux Balance Approach." *Environmental Microbiology*, **4(3)**, 133–140.

Soetaert K, van Oevelen D (2009a). *LIM: Linear Inverse Model examples and solution methods.* R package version 1.4.

Soetaert K, van Oevelen D (2009b). *Package LIM, implementing linear inverse models in R*. R package vignette.

van Oevelen D, van den Meersche K, Meysman F, Soetaert K, Middelburg JJ, Vezina AF (2009). "Quantifying Food Web Flows Using Linear Inverse Models." *Ecosystems*. doi:10.1007/s10021-009-9297-6. URL http://www.springerlink.com/content/4q6h4011511731m5/fulltext.pdf.

Affiliation:

Karline Soetaert
Centre for Estuarine and Marine Ecology (CEME)
Netherlands Institute of Ecology (NIOO)
4401 NT Yerseke, Netherlands E-mail: k.soetaert@nioo.knaw.nl
URL: http://www.nioo.knaw.nl/users/ksoetaert