Concentrations of six chemical components $z_1, \dots, z_6$ during biodiesel production reaction can be described with the following system of ordinary differential equations. We want to fit this model to experimental data about concentrations of the components in time.

$$\begin{cases} \frac{dz_1}{dt} = -k_1 z_3 z_1 + k_2 z_4 z_6 - k_3 z_4 z_1 + k_4 z_5 z_6 - k_5 z_5 z_1 + k_6 z_2 z_6 \\ \frac{dz_2}{dt} = k_5 z_5 z_1 - k_6 z_2 z_6 \\ \frac{dz_3}{dt} = -k_1 z_3 z_1 + k_2 z_4 z_6 \\ \frac{dz_4}{dt} = k_1 z_3 z_1 - k_2 z_4 z_6 - k_3 z_4 z_1 + k_4 z_5 z_6 \\ \frac{dz_5}{dt} = k_3 z_4 z_1 - k_4 z_5 z_6 - k_5 z_5 z_1 + k_6 z_2 z_6 \\ \frac{dz_6}{dt} = k_1 z_3 z_1 - k_2 z_4 z_6 + k_3 z_4 z_1 - k_4 z_5 z_6 + k_5 z_5 z_1 - k_6 z_2 z_6 \end{cases} \qquad \mathbf{1.1}$$

Regression task: find $k_1, \dots, k_6$ which for given initial condition $z_{1,t_0}, \dots, z_{6,t_0}$ best fit the experimental data $z_{1,t}, \dots, z_{6,t}$ for $t = 2, 4, 6, \dots, 70$. Variable $z_{i,t}$ denotes the concentration of component i measured at time t, whereas $z_i(t)$ for $i = 1, \dots, 6$ is a solution to the system 1.1.

Error measure for one component and time t

$$\epsilon_{i,t} = |z_{i,t} - z_i(t)| \qquad \mathbf{1.2}$$

We want to aggregate these errors using Lp norm:

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p\right)^{\frac{1}{p}} \qquad \mathbf{13}$$

We have a two-step error aggregation, first for each component over time with norm Lp and then over components with norm Lq

$$\epsilon_{fit} = \left\| \|\epsilon_{1,t}\|_p \; \|\epsilon_{2,t}\|_p \; \|\epsilon_{3,t}\|_p \; \|\epsilon_{4,t}\|_p \; \|\epsilon_{5,t}\|_p \; \|\epsilon_{6,t}\|_p \right\|_q \qquad \mathbf{1.4}$$

To make the solutions more robust to inaccurate (or sometimes incorrect) data we try to use $p < 2$, namely $p = 0.25$, $p = 0.5$ and $p = 1$.

Concentrations of some components differ in about an order of magnitude. Maybe we could measure error for normalized (studentized) data? On the other hand performing the second norming in eq. 1.4 with $q < 1$ may have similar effect?

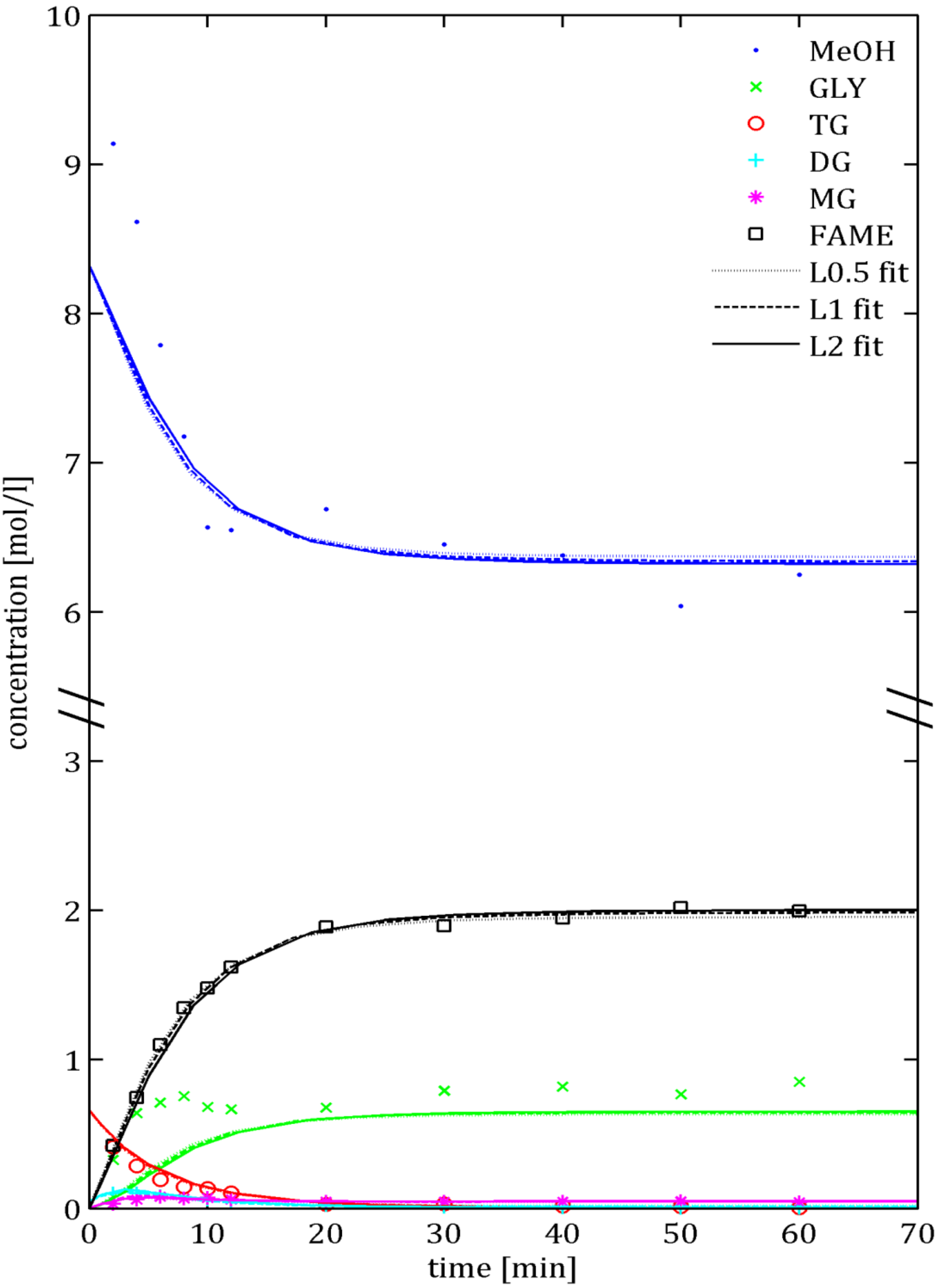
$$\epsilon_{i,t} = \frac{|z_{i,t} - z_i(t)|}{\mathrm{std}_{s \in T}(z_{i,s})} \qquad \mathbf{1.5}$$

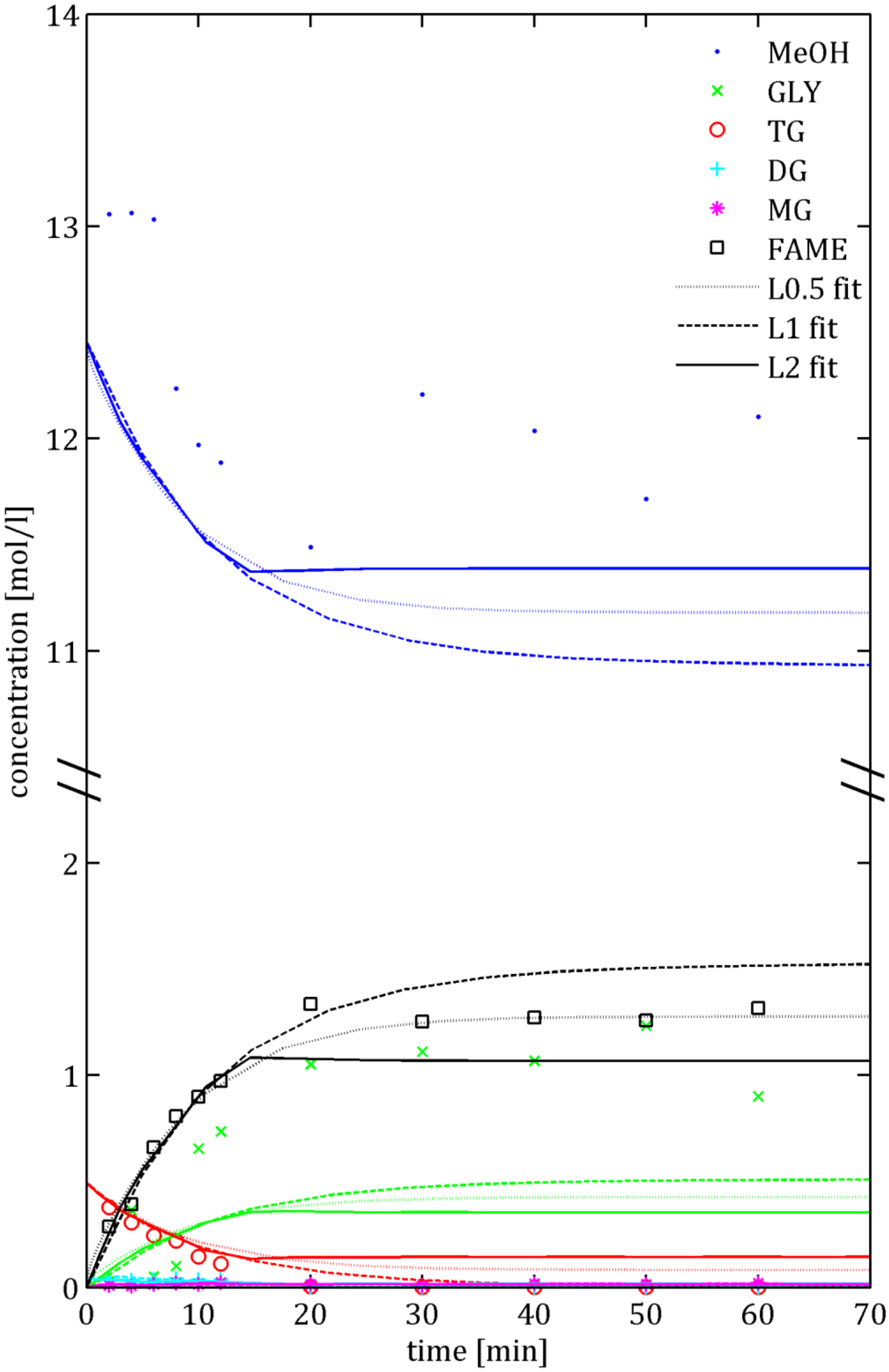
Problems to consult:

1. How would you comment on using the Lp norm with small p for regression?
2. What is the best way to proceed with aggregation among components?
    a. First normalize (or studentize) the error data and use mean squared error $q = 2$
    b. Use normalization with $q < 1$ to increase the relative importance of small errors
    c. Use some other method?

Appendix

Below, please find two sets of experimental data, in the first data is more-less correct, while in the second we have excess of component GLY denoted with green crosses (it should be at 1/3 of the component FAME). These plots were created with $q = p = 0.5, 1, 2$.





MeOH
GLY
TG
DG
MG
FAME
L0.5 fit
L1 fit
L2 fit
concentration [mol/l]
time [min]
0
1
2
11
12
13
14
0
10
20
30
40
50
60
70