BT5420 - COMPUTATIONAL SYSTEMS BIOLOGY

Assignment 3

Submitted by Sahana (BE17B038)

CONTENT

Files/Codes attached –

1. **ode_analyser.m**
   This function stores the values for all variables. It also contains the rate equations of all the reactions and the differential equation corresponding to all the metabolites.
2. **compute_err.m**
   a. Input to this function – double array (75*9) containing the data points and the parameter values of p1, p2, p3.
   b. **ode23s** is used as the ODE solver to predict data at the same time points as the given data.
   c. The predicted and the measured data are compared via a cost/error function to analyse the fit of the input set of parameters to the given system.
   d. Error function is as follows -
   E = sum ((predicted value – measured value) ^2) across all the data points.
   e. Output of this function – E: error value, and x_pred: predicted data for metabolites' concentration.
3. **multiple_compute_err.m**
   a. Input to this function – two double arrays (75*9) containing the data points and the parameter values of p1, p2, p3.
   b. **ode23s** is used as the ODE solver to predict data at the same points as the given data.
   c. The predicted data is compared with both the measured data sets via a cost/error function to analyse the fit of the input set of parameters to the given system.
   d. Error function is as follows -
   E = sum ((predicted value – measured value) ^2) across all the data points with **both** the measured data sets.
   e. Output of this function – E: error value, and x_pred: predicted data for metabolites' concentration.
4. **fminsearchbnd.m**
   a. This is a publicly available add-on program in MATLAB, which implements the vanilla '**fminsearch'** algorithm, and takes Lower and Upper bounds for parameter values as additional input arguments.
   b. Input to this function – [Fun, X0, LB, UB]
      i. Fun - Function that needs to be minimised
      ii. X0 - Initial conditions for the parameters
      iii. LB – Lower bound for the parameters
      iv. UB - Upper bound for the parameters
   c. Output to this function – Estimated parameter values and error value associated with it.
5. **hw3_show.m**
   a. Input to this program - double array (75*9) containing the measured data
   b. The **fminsearchbnd** function is called with necessary arguments. The output of this function, i.e., the parameters (X) and error value (FVAL) are stored in two variables and they are displayed. The objective function that is passed as an input argument is *compute_err*.

c. The **compute_err** function is called with input arguments as the measured data and the parameter set (X). The output of this function, i.e., the error value (E) and predicted data (x_values) are stored in two different variables.
d. The first column in the measured data is the time points. This is stored in a variable (t).
e. The metabolite concentration (8 columns of x_values) is plotted with respect to time (t), in the same figure, and the trend in variation of concentration is analysed.

6. **hw3_multiple_show.m**
    a. Input to this program – two double arrays (75*9) containing the measured data
    b. The **fminsearchbnd** function is called with necessary arguments. The output of this function, i.e., the parameters (X) and error value (FVAL) are stored in two variables and they are displayed. The objective function that is passed as an input argument is *multiple_compute_err*.
    c. The **multiple_compute_err** function is called with input arguments as the measured data and the parameter set (X). The output of this function, i.e., the error value (E) and predicted data (x_values) are stored in two different variables.
    d. The first column in the measured data is the time points. This is stored in a variable (t).
    e. The metabolite concentration (8 columns of x_values) is plotted with respect to time (t), in the same figure, and the trend in variation of concentration is analysed.

Remarks –

1. **Choice of error function –**
    a. **E = sum ((predicted value – measured value) ^2)**
    b. There are other plausible error functions where we can normalize the above equation with the measure value. But normalizing will mean that we are discounting few wrong data points if they have large value of measured data, i.e., x (i, j). This needn't be done.
    c. Hence, the error function can be the square of the difference between predicted and measured data points, which is what I've used.
2. Why **fminsearchbnd**?
    a. **fminsearch** is based on Nelder-Mead simplex search algorithm which is the best for solving multidimensional unconstrained optimization. But this algorithm, as mentioned is unconstrained, hence it is possible to get negative, large numbers for our parameters, because that set would yield the minimum value for the cost function.
    b. An intuitive alternative would be to impose boundary constraints for the parameters that we are finding. This is exactly what fminsearchbnd does, and is hence used as the optimizing algorithm.
3. **Initial conditions for optimization algorithm–**
    a. The given system is very sensitive to initial conditions, i.e., even the slightest change in initial conditions, yield very different results for the parameter set.

b. In a trial and error fashion, the initial conditions for all the 6 sets of observation was found.
    i. I iterated the initial conditions as [i; i; i] where i goes from 0 to 10.
    ii. For cases where we are optimizing with single set of measured data, and for both the variable sets (Set 1 and 2), the obtained error value was comparatively lesser with initial conditions = [0; 0; 0].
    iii. For cases where we are optimizing with two sets of measured data, and for both the variable sets (Set 1 and 2), the obtained error value was comparatively lesser with initial conditions = [1; 1; 1].

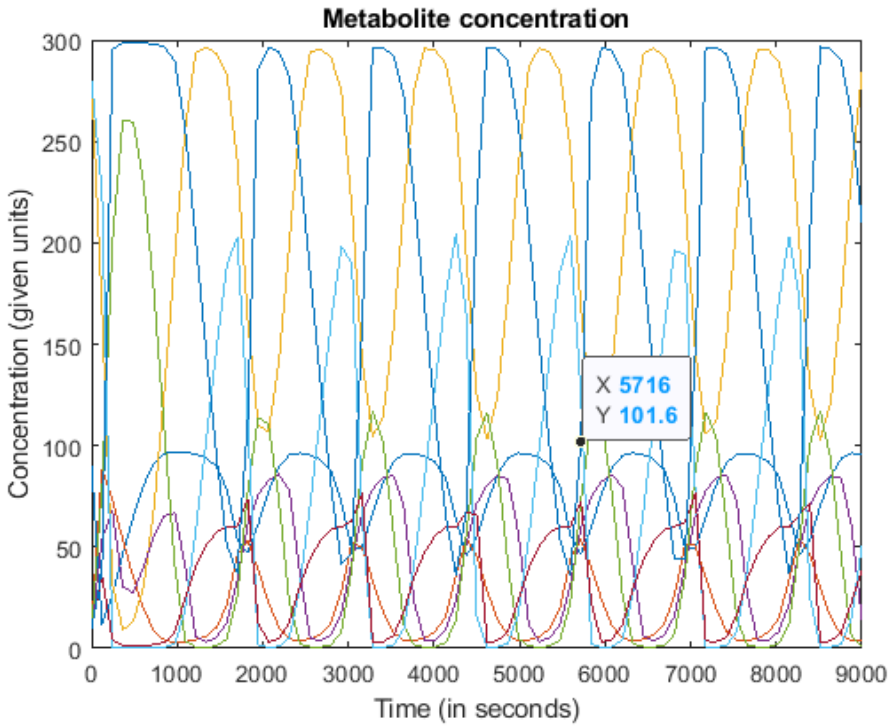
4. **Moiety Conservation relation** –
    a. The question statement gives three conservation relations for the metabolites.
    b. Although, not specified explicitly, this has been incorporated and here's how.
    c. The initial values for these metabolites (x0(:) – mentioned in compute_err and multiple_compute_err) follow the above stated relations.
    d. The differential equation corresponding to all metabolites (in ode_analyser.m) is such that the metabolites are conserved. For example -
        i. d[MKKK]/dt + d[MKKK-P]/dt = 0 (since, their combined total value is always constant)
        ii. d[MKK]/dt + d[MKK-P]/dt + d[MKK-PP]/dt = 0
        iii. d[MAPK]/dt + d[MAPK-P]/dt = 0
5. **The difference in initial values of the metabolites** (at t=0) between the measured value and the initial conditions that we set in compute_err.m and multiple_compute_err.m is counted as a part of the error in the measured data.
6. **Choice of ODE Solver**
    a. The given system is comprised of differential equations where the rate of change of metabolites is unknown. Hence, it is better to not assume anything about our system and hence use a stiff solver for the problem.
    b. **Ode23s** has been used because it is known to be efficient with crude error tolerances.
    c. But the answer (parameter set) does not vary much with ode15s (the other commonly used solver for stiff systems).

Observations –

Dataset1, Set 1 parameters

Initial condition for the optimizing function = [0; 0; 0]

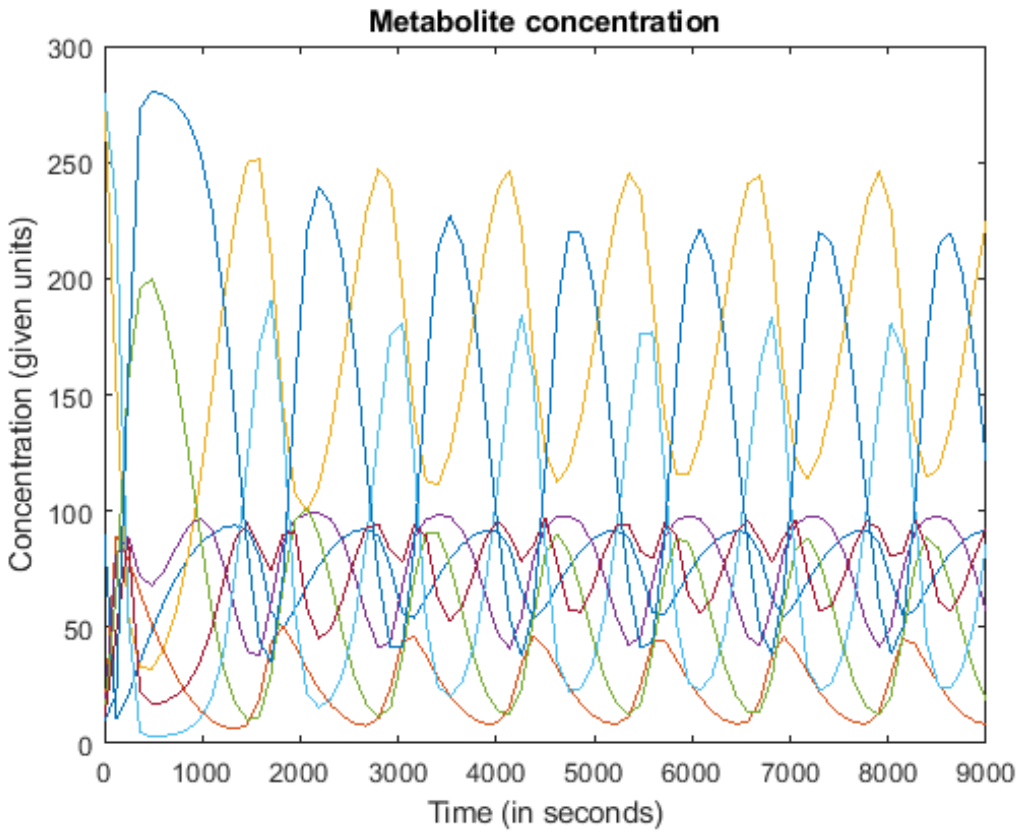
```
>> hw3_show(A)

Exiting: Maximum number of function evaluations has been exceeded
         - increase MaxFunEvals option.
         Current function value: 61190.247213

Parameter values:
    0.0249
    0.7495
    0.2516

Value of cost function for the above parameter set:
   6.1190e+04

Displaying the variation in the metabolites:
```



Dataset1, Set 2 parameters

Initial condition for the optimizing function = [0; 0; 0]

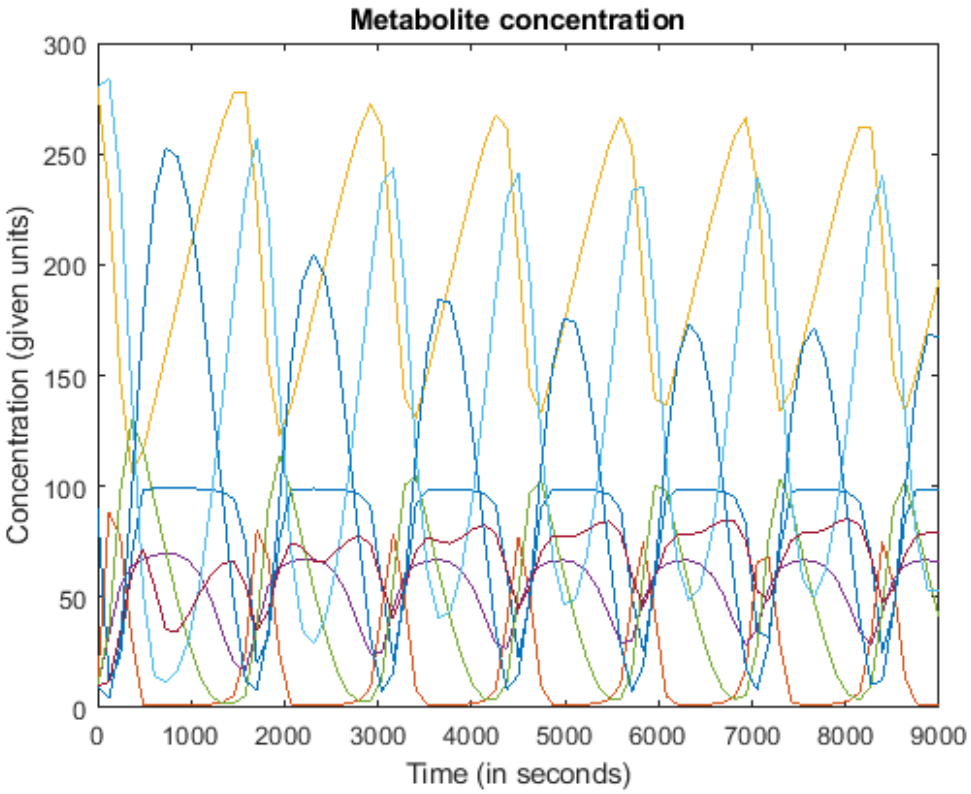
```
>> hw3_show(A)
Parameter values:
    0.0322
    1.0908
    0.1966

Value of cost function for the above parameter set:
   6.3091e+05

Displaying the variation in the metabolites:
```



Dataset2, Set 1 parameters

Initial condition for the optimizing function = [0; 0; 0]

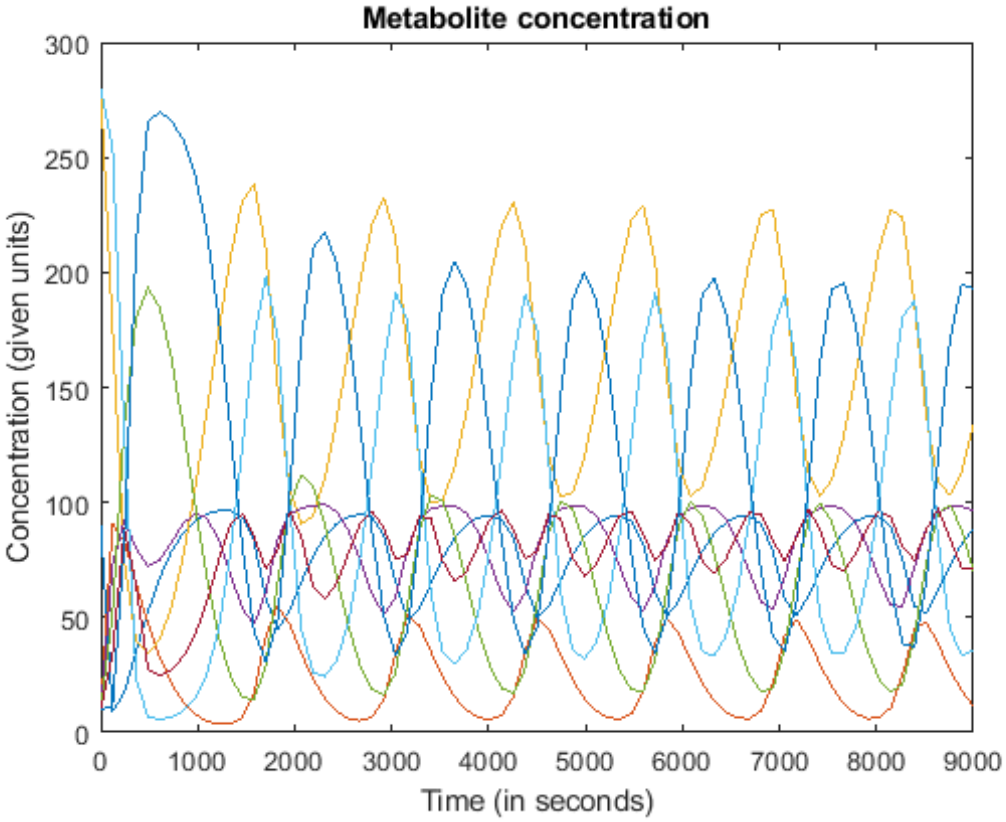
```
>> hw3_show(B)
Parameter values:
    0.0107
    0.2330
    0.7537

Value of cost function for the above parameter set:
   6.0929e+05

Displaying the variation in the metabolites:
```



Dataset2, Set 2 parameters

Initial condition for the optimizing function = [0; 0; 0]

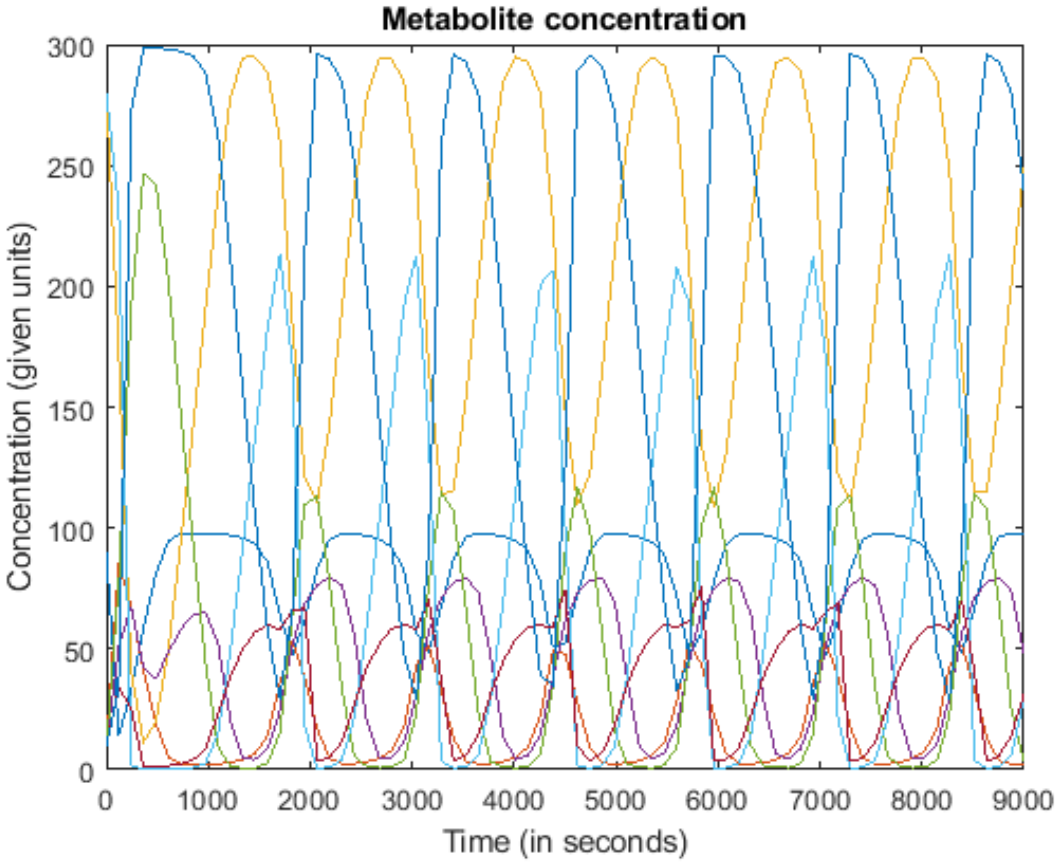
```
>> hw3_show(B)
Parameter values:
    0.0246
    0.7423
    0.2531

Value of cost function for the above parameter set:
   5.2780e+04

Displaying the variation in the metabolites:
```



Dataset 1 U 2, Set 1 parameters

Initial condition for the optimizing function = [1; 1; 1]

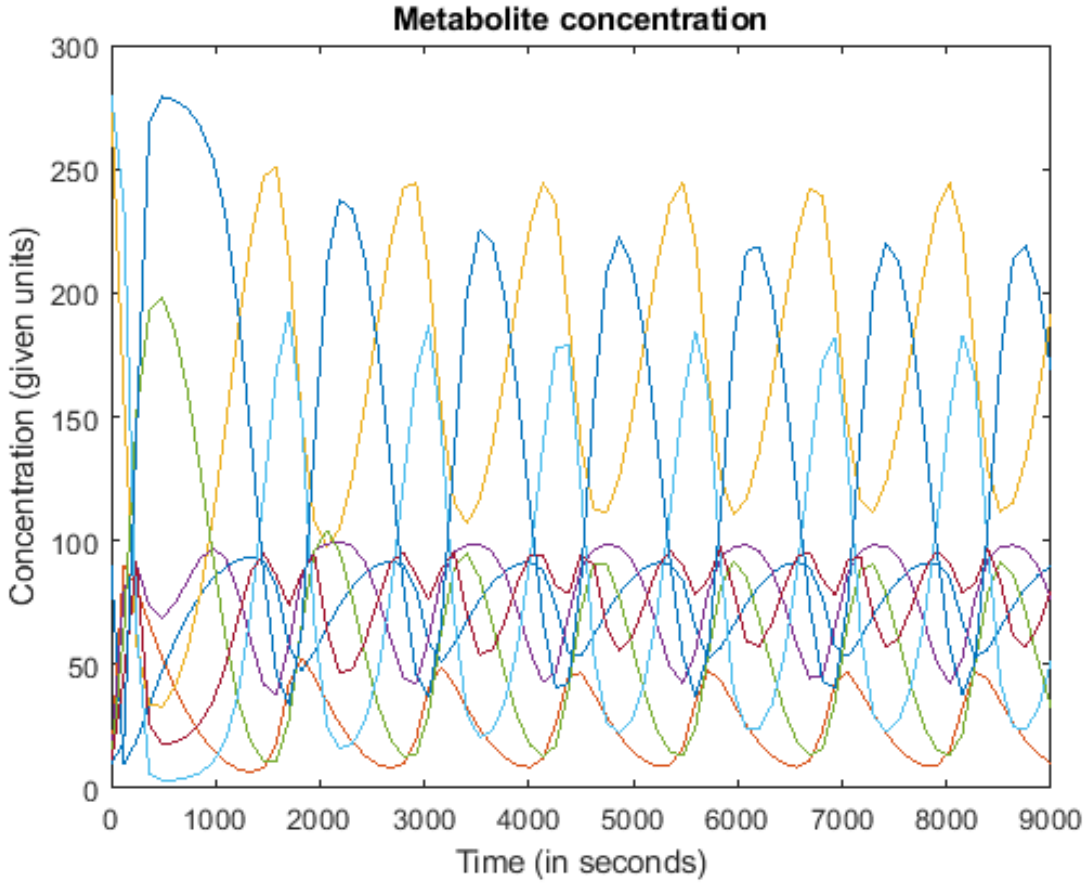
```
>> hw3_multiple_show(A,B)

Exiting: Maximum number of function evaluations has been exceeded
         - increase MaxFunEvals option.
         Current function value: 1303516.462082

Parameter values:
    0.0226
    0.5470
    0.3164

Value of cost function for the above parameter set:
   1.3035e+06

Displaying the variation in the metabolites:
```



Dataset 1 U 2, Set 2 parameters

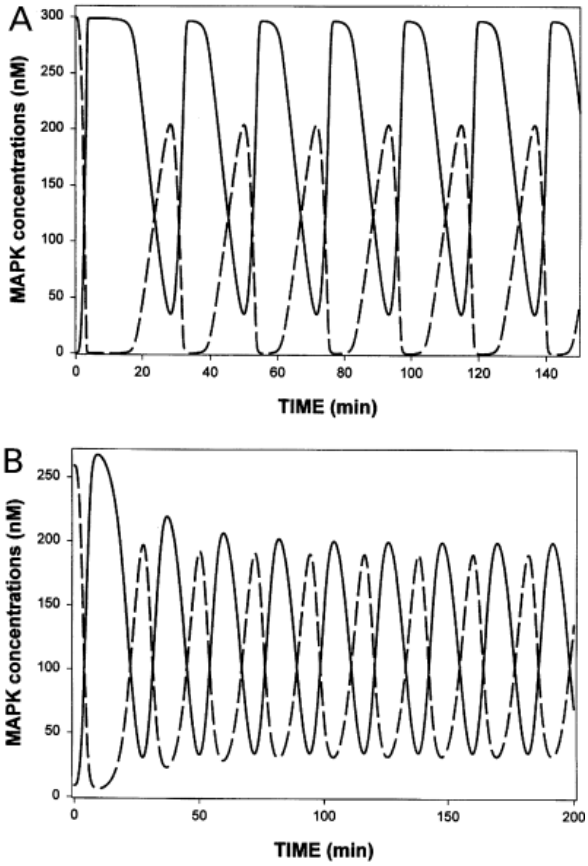
Initial condition for the optimizing function = [1; 1; 1]

```
>> hw3_multiple_show(A,B)
Parameter values:
    0.0310
    1.0790
    0.1939

Value of cost function for the above parameter set:
   9.6896e+05

Displaying the variation in the metabolites:
```



Discussion –

1. **Oscillation in the metabolite concentration** -
    a. The above system is in fact a model that explains the following paper - Kholodenko, B.N. (2000), Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. European Journal of Biochemistry, 267: 1583-1588. doi:10.1046/j.1432-1327.2000.01197.x
    b. It is observed that the metabolites undergo oscillations in their concentration.



    c. The phosphorylation cascade has a positive and negative feedback loop.
        i. A phosphorylation cascade with a strong positive feedback can have three different steady states, stable 'off' and stable 'on' states with low and high phosphorylation levels, respectively, separated by an unstable state that corresponds to a threshold level.
        ii. Sustained oscillations are an emerging property of ultrasensitive cascades with negative feedback, as there is always a range of kinetic constants in which oscillatory behaviour is observed.
2. The system is very sensitive to changes in the method of calculating error, initial values for the parameter sets and also the choice of algorithm used to solve the optimization problem. Hence, one must be very careful while picking the above variables.

Result –

Error = Sum of squared error'
error' = difference between predicted and measured data points

S. No.	Dataset	Parameter set	Estimated p1	Estimated p2	Estimated p3	Error value
1	1	1	0.0249	0.7495	0.2516	6.1190e+04
2	1	2	0.0322	1.0908	0.1966	6.3091e+05
3	2	1	0.0107	0.2330	0.7537	6.0929e+05
4	2	2	0.0246	0.7423	0.2531	5.2780e+04
5	1 U 2	1	0.0226	0.5470	0.3164	1.3035e+06
6	1 U 2	2	0.0310	1.0790	0.1939	9.6896e+05

- The error values are pretty high, but are comparable with other datasets and parameter values, and hence should not be a problem as we adopt a comparative method for analysing the observations. The reasoning for the choice of error function is mentioned in the Remarks section.
- Case #1, #4 are given comparatively good fits by the algorithm than for the remaining cases.