**Table of Indices**

|  |  |  |
| --- | --- | --- |
| **Index** | **Description** | **Count** |
| *t* | ~~Index used for values tracked over each time step~~ | Total simulation time/time step size |
| *j* | ~~Index used for values related to different experiments~~ | Total experiments analyzed |
| *k* | ~~Index used for the number of growth phenotypes analyzed. Note, every species needs to have one phenotype dedicated to the stationary phase, which should have no growth and no metabolite consumption.~~ | Total growth phenotypes |
| *i* | Index used for each metabolite tracked in the simulation | Total metabolites consumed or excreted by all growth phenotypes |

**Table of Problem Parameters:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Index, colDescription** | **Obtained from** |
| df.at[index, col]  (rfpt,j) | ~~Measured RFP abundance at time~~ *~~t~~* ~~for experiment~~ *~~j~~* | Read for Jeff’s data |
| species\_phenotypes\_bool\_df  (rfpsk) | ~~Set to one if strain~~ *~~k~~* ~~is~~ *~~Pseudomonas~~* ~~and zero otherwise~~ | Read from phenotype table |
| timestep\_s  (dt) | ~~Size of time step used in simulation~~ | Optimize value based on simulation performance |
| growth\_stoich  () | ~~Stoichiometry for interaction of strain~~ *~~k~~* ~~with metabolite~~ *~~i~~* | Read from phenotype table |
| v  () | ~~Coefficient controlling how quickly strain k grows at time step t in experiment j. Initially, I envision this being a constant value for each distinct strain k, perhaps varying per experiment, but perhaps not. This is equivalent to assuming nutrient consumption rates are in the saturated regime of Michaelis–Menten (MM) kinetics and independent of nutrient concentration. I envision us optimizing these initial fixed constants to maximize the overall data fit by simply varying the values iteratively until the objective function stops improving. If the overall data fit is still poor at this point, then I envision us introducing km values from MM kinetics and varying the parameter values over time based on the substrate concentration from the previous simulation. So we simulate with fixed MM kinetics, compute the optimal fixed~~ ~~values, which will give us a simulation with varying metabolic concentration values. We then use those values to compute new varying values according to the MM expression: . We then iteratively repeat the simulation recomputing until the concentration profiles settle. We also iterate to determine ideal values for vmax and km in the above equation.~~ | Optimized iteratively to maximize the fit of the predict biomass profiles with the experimental data. |
| cvct  () | ~~Coefficient for the minimization of phenotype conversion to the stationary phase. We will need to carefully tune this so phenotype conversion is minimized without hurting the data fitting.~~ | Tuned iteratively in simulations. |
| cvcf  () | ~~Coefficient for the minimization of phenotype conversion from the stationary phase. We will need to carefully tune this so phenotype conversion is minimized without hurting the data fitting.~~ | Tuned iteratively in simulations. |
| bcv  () | ~~This is the highest fraction of biomass for a given strain that can change phenotypes in a single time step. If we set this to “1”, we effectively say there is no limit, and that’s probably the initial value we should use for this parameter.~~ | Tuned iteratively in simulations. |
| cvmin  () | ~~This is the lowest value the limit on phenotype conversion goes, ensuring that ALL biomass can eventually convert even if we use a~~ *~~bcv~~* ~~value that is less than “1”.~~ | Tuned iteratively in simulations. |

**Variables:**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Name** | **Description** | **Type** | **LB** | **UB** |
| {spcs}\_\_conversion (rfbc, gfbc, odc) | ~~Coefficient for converting species signal/OD values into biomass abundances~~ | Cont | 0 | 1000 |
| {spcs}\_\_bio  (rfpb, gfpb, odb) | ~~Biomass computed from experimental signals for each species at time~~ *~~t~~* ~~and in experiment~~ *~~j~~* | Cont | 0 | 1000 |
| {spcs}\_\_diff  (rfpe, gfpe, ode) | ~~Difference between predicted and measured biomass for each species/OD at time~~ *~~t~~* ~~and in experiment~~ *~~j~~* | Cont | -100 | 100 |
| b\_{strain}  (b) | ~~Model-based biomass abundance of strain~~ *~~k~~* ~~at time~~ *~~t~~* ~~for experiment~~ *~~j~~* | Cont | 0 | 1000 |
| c\_{met}  (c) | ~~Concentration of metabolite~~ *~~I~~* ~~at time~~ *~~t~~* ~~for experiment~~ *~~j~~* | Cont | 0 | 1000 |
| g\_{strain}  (g) | ~~Growth rate of strain~~ *~~k~~* ~~at time~~ *~~t~~* ~~for experiment~~ *~~j~~* | Cont | 0 | 1000 |
| cvt\_{strain}  (cvt) | ~~Rate of conversion of strain k to the stationary phase for each species. Of course, the stationary phase strains will not have this variable. Note, to capture the concept of lag associated with switching phenotypes, strains will never go directly from one phenotype to another. They will always convert first to the stationary phase phenotype, then to a different phenotype in the following time step.~~ | Cont | 0 | 100 |
| cvf\_{strain}  (cvf) | ~~Rate of conversion of strain k from the stationary phase for each species. Same notes as above variable.~~ | Cont | 0 | 100 |

**Constraints:**

|  |  |  |
| --- | --- | --- |
| **Name** | **Equation** | **Description** |
| {spcs}\_\_bioc(gfpbct,j) |  | ~~Setting value of gfp-based biomass abundance based on gfp signal~~ |
| {spcs}\_\_diffc  (gfpect,j) |  | ~~Setting value for the different between the gfp-based biomass abundance and the model predicted gfp-biomass abundance~~ |
| *odect,j* |  | ~~Setting value for the different between the OD-based biomass abundance and the model predicted OD-biomass abundance~~ |
| *dbct,j,k* | ~~Nonstationary:~~    ~~Stationary:~~ | Iterating biomass value for each strain at each time point based on strain growth rate and transition of strains to or from other phenotypes (*cv* term). A different equation is used depending on whether the strain is stationary or nonstationary (which switches the sign on the terms that move strains between phases. |
| dcc\_{met}  (dcct,j,i) |  | Iterating metabolite concentration for each metabolite at each time point based on strain growth rate |
| gc\_{strain}  (gct,j,k) |  | Setting the growth rate of each strain at each time step in each experiment. Note, this expression does not directly accommodate for the fact that growth must stop when substrate runs out, and substate concentrations can never be negative. However, because this formulation solves for all time steps simultaneously and has the capacity to transition some biomass to stationary phase at each time point, it should be possible to perfectly tune biomass abundance to avoid overconsuming substrates. |
| cvc\_{strain}  (cvct,j,k) |  | Setting an upper limit on what fraction of biomass can convert to the stationary phase in a single time step. It’s unclear if there needs to be such a limit, but it may make sense. Still, there needs to be a minimum value because otherwise, all the biomass could never convert completely and we would have problems overconsuming substrates. |

**Objective:**

|  |  |  |
| --- | --- | --- |
| **Sense** | **Equation** | **Description** |
| Minimize |  | Minimize sum of square variation between model-predicted biomass abundances and experimentally measured biomass abundances while penalizing transition between phenotypes to avoid this mechanism being abused to overfit data. We will need to carefully tune the *cvc* coefficient in that term to ensure it is balanced properly with the data fitting. |