# ODE model tranfer between oncogenic cell lines.

## Introduction to project

## Required equipment/software to continue:

A computer capable of running COPASI.

COPASI (<http://copasi.org/>) (explorative model building)

Python 2.7 (for using pycotools)

Pycotools (<https://pycotools.readthedocs.io/en/copasiversion21/QuickStart.html>) (parameter estimation, identifiability and validation analysis)

Cell designer ([http://www.celldesigner.org](http://www.celldesigner.org/), model visualization)

Data of MCF7, T47D and ZR-75-1 cell lines was acquired by Lea Timpen under supervision of Ulrike Bosch and Karen van Eunen. For initial model setup and parameterization MCF7 was used as this is the most commonly used ER+ breast cancer cell type.

For a quick update on mTOR one can read the following reviews:

Saxton and Sabatini ([28283069](https://www.ncbi.nlm.nih.gov/pubmed/28283069))

Razquin and Thedieck ([28698309](https://www.ncbi.nlm.nih.gov/pubmed/?term=differntial+control+of+ageing+and+lifespan+by+isoforms+and+splice+variants+across+the+mtor+network))

Document and data locations:

Y:\staff\LowCost\UMC\KINDGNK\E4\Kathrin\SG\ODE\_models

This location contains all data (non-normalized/normalized/transformed for pycotools). It also contains the cell designer xml files and figures and the COPASI files for the different models with their results.

Advice:

Transfer the model into data2dynamics as it offers more control and flexibility over model parameterization and regulation. It is potentially easier to expend the established model with ER+ breast cancer pathways. Steps model can be finished using COPASI however for expanding the model with for example estrogen receptor cascade data2dynamics would be more suited.

Appendix:

1. Table describing changes between models

## Goal:

To create an ODE model of the mTORC1/2 signaling pathways that can be transferred between cell lines.

## Current progress and decisions:

The model of Della Pezze et al([22457331](https://www.ncbi.nlm.nih.gov/pubmed/22457331)) has been used as a basis to develop the current model as shown in figure 1. There are several simplifications in the TSC1/2, AMPK and IRS signaling. For transferring the model between cell lines one is referred to the paper of Adlung et al([28123004](https://www.ncbi.nlm.nih.gov/pubmed/?term=Protein+abundance+of+AKT+and+ERK+pathway+components+governs+cell+type‐specific+regulation+of+proliferation)), which transferred a model between non-oncogenic cell lines. To reduce the chance of unidentifiable parameters the model was shrunk, the TSC1/2, AMPK and IRS module where selected to be simplified as there was little or no experimental data present. 4EBP was added as an output as experimental data was available. Initial model topologies where based on species and reaction in and outputs, however as simulations showed that this did not work the model was changed to incorporate global quantities (these are two different ways of writing a model in COPASI). Due to the inability of the experimental techniques used to discriminate between the single and double phosphorylated sites it is required to calculate the balance between the too. As it is unclear what amount of signal belongs to a single or double phosphorylated amount it should be calculated. To do this total phosphorylation was expressed as the sum of the single and double phosphorylated state. An example equation is given below, In this way the algorithm itself figures out the correct balance between the two quantitates.

The first step was the estimation of initial values of unknown model species given the known initial concentrations. This was performed by only selecting the transient concentrations (under species) within COPASI and running a genetic algorithm (GA) with 500 generations having a population size of 75 followed by a Hooke and Jeeves (HJ) local optimizer with 1000 iterations and tolerance of 1e-10. The resulting parameters where kept and the next step was testing various global search algorithms to see if the current topology could generate correct fits of the data. Some examples of these fits are given in figure 2. The result from figure 2 where obtained using a GA with 500 generations having a population size of 75, range 10e-6, 10e4.

From the initial estimates it can be seen that the model using global quantities predicts most phosphorylation time course data (RSS = 0.109). However, for most the model does not capture the change point, this could be due to an inherent topology issue, due to the fact that no global optima was reached or that the equations do not allow for enough flexibility. It can also be due to the lack of experimental data in the lower time points. Before addressing this issue the search space was reduced from 1e-6 to 1e4 to 1e-4 to 1e3. The reason is that the range are still biologically relevant but the reduction in search space should reduce the chance of non-identifiability after a global optima has been selected. The range was selected by systematically reducing the search space until the final RSS best value increased (n=3). To increase the chance of finding the global optima, 638 models using the topology of figure 1 where run with a global parameter search. The method used GA with 500 generations having a population size of 75. The estimates improved little and mainly drew the predicated values close to those after the change point.

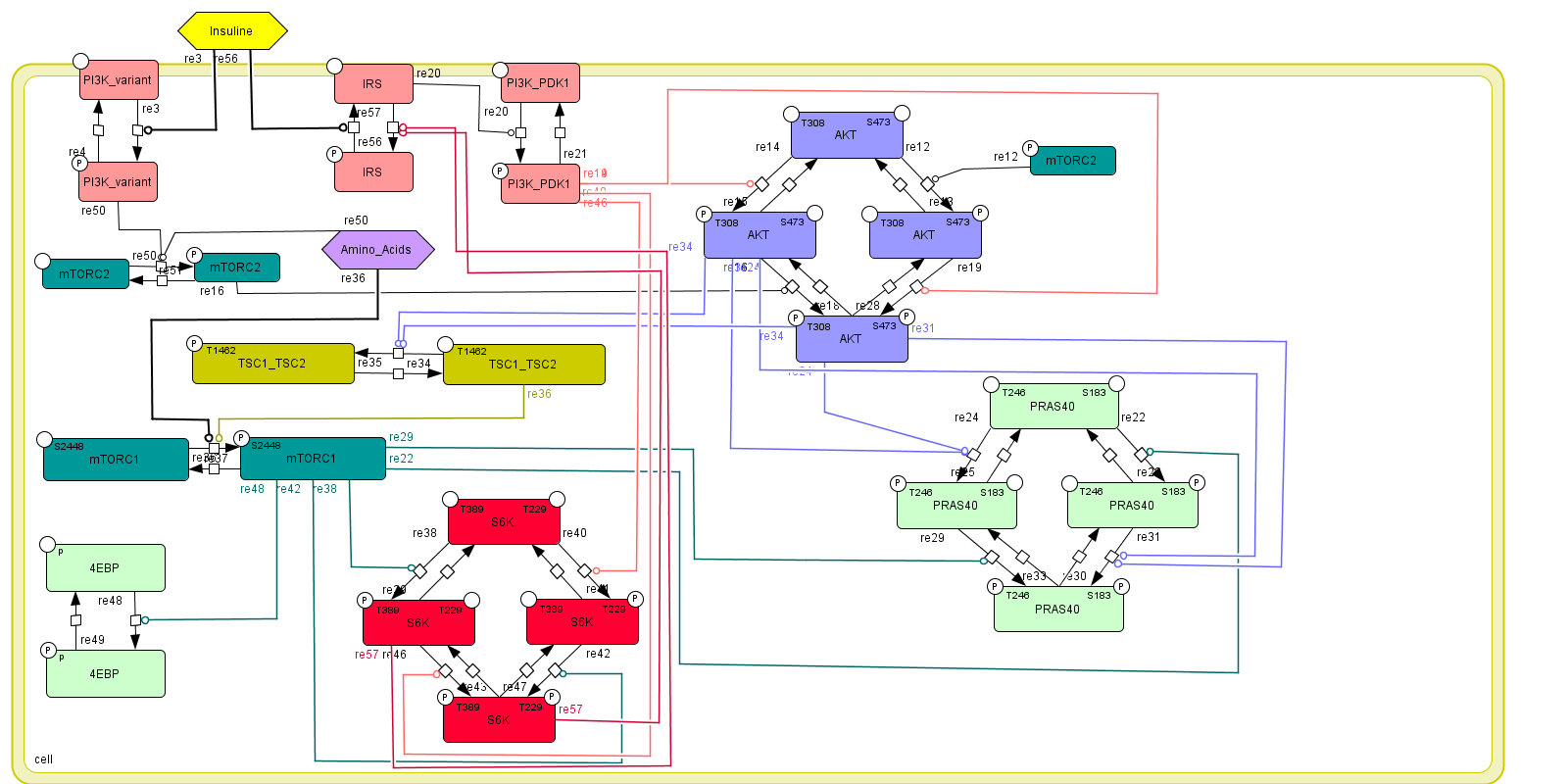
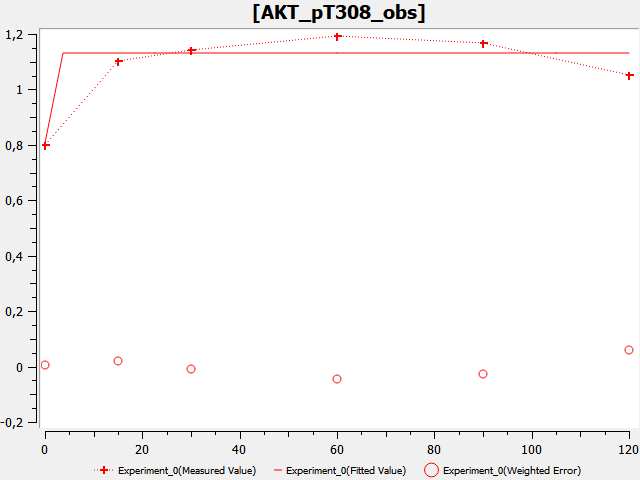
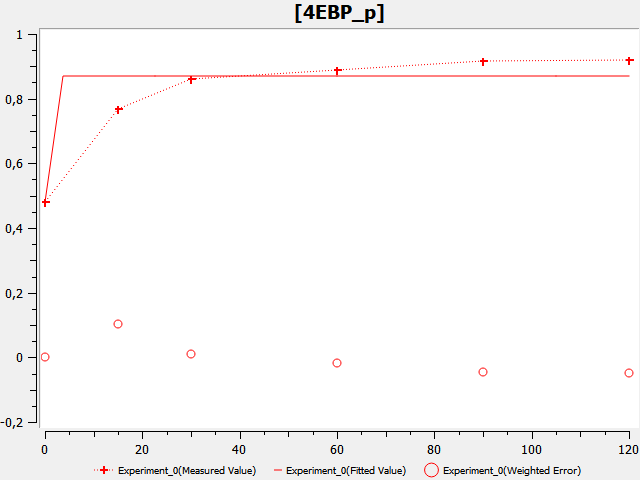
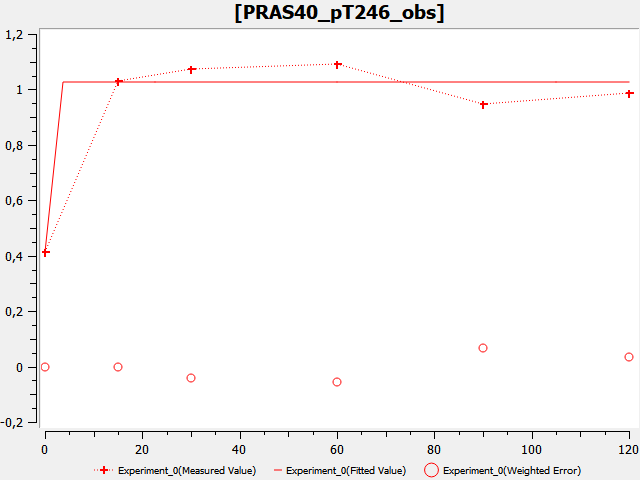


Figure 1. Used model topology





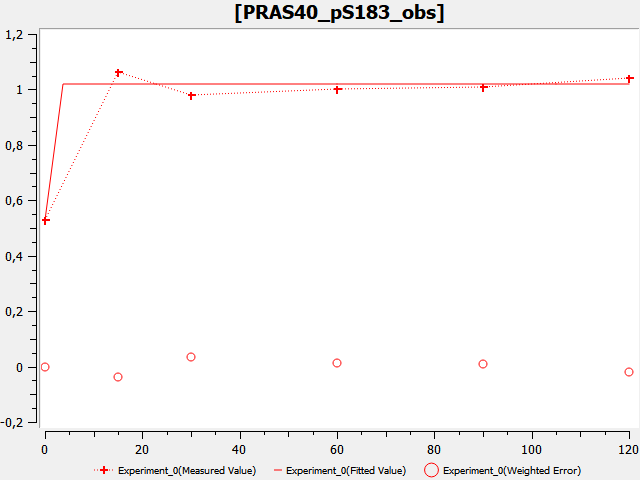
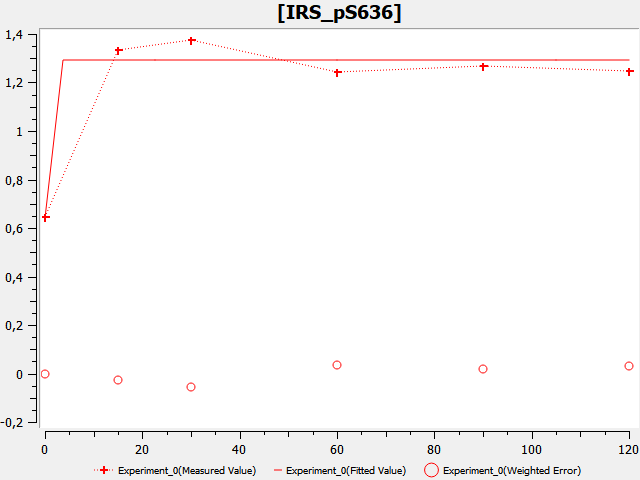
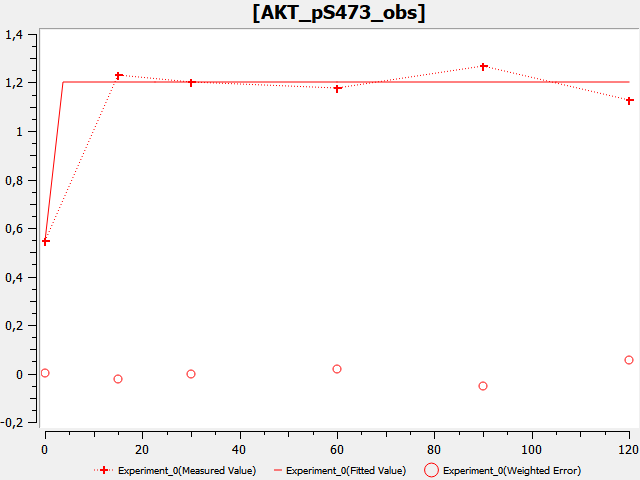


Figure 2. Initial estimates using mass-action equations

As apparently the mass action equations are unable to capture the trends of the data these where exchanged for Michealis-Menten equations with 1 or 2 modifiers are shown below. Where Km represents the too estimate parameter value and kcat was kept at 1 for all reactions. Ignoring kcat was done to keep the amount of estimated parameters the same as the previous parameter estimation, this mean that the chance of non-identifiability is not increased. These equations are different than the ones found in the previous papers however, these experiments contained many more time measurements in the lower time region. Hence these models did not need regulation to prevent overfitting in the starting time region.

The new initial estimates for model using Michealis-Menten kinetics can be seen in figure 3. The parameter estimation was performed the same way as with the mass-action equations. The RSS was a bit higher after the global search. RSS = 0.233. However the trends seem to be better captured. Therefor PS models with an iteration limit of 2500 and a swarm size of 50 where run. The resulting log-likelihood plot is shown in figure 4. It is clear that these plots better capture the change point and seems to better predict the biological change. Parameter boxplots are shown in figure 5. In addition estimations were run for the other 2 datasets of T47D and ZR-75-1, which also gave reasonable fits. From the distributions seen in figure 6 4EBP, 4EBPp, AKT and S6K seem to have narrow distributions indicating that they could be identifiable. However, there are several parameters that have large parameter distributions. This indicates that they could be non-identifiable (note that this could change when identifiable parameters are fixed). As the result seemed promising and hence identifiability analysis was performed.

The model with the lowest RSS was selected for further optimization using the Hooke and Jeeves algorithm with iteration limit of 1000 and a tolerance of 1e-10. Identifiability analysis was performed using pycotools with an interval of 20. Note that these figures are shown only for the final model topology but that the results of previous models can be found in their respective folders.

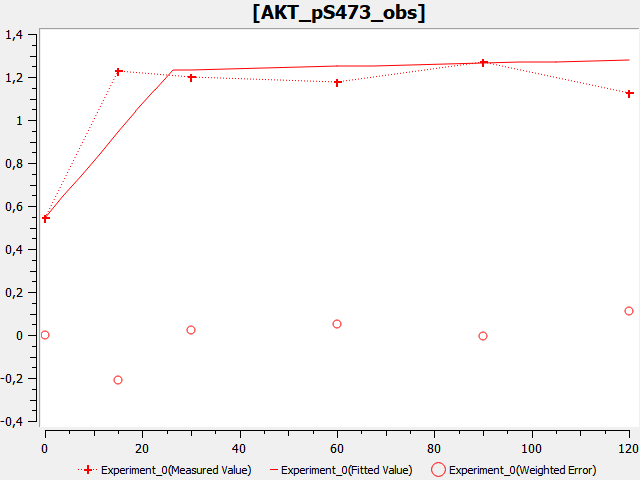
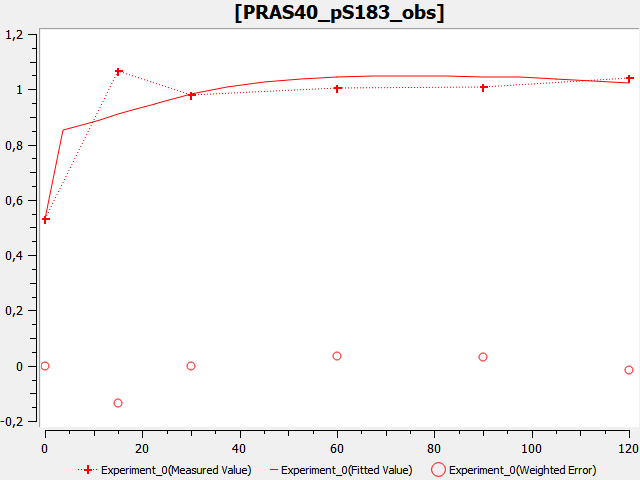
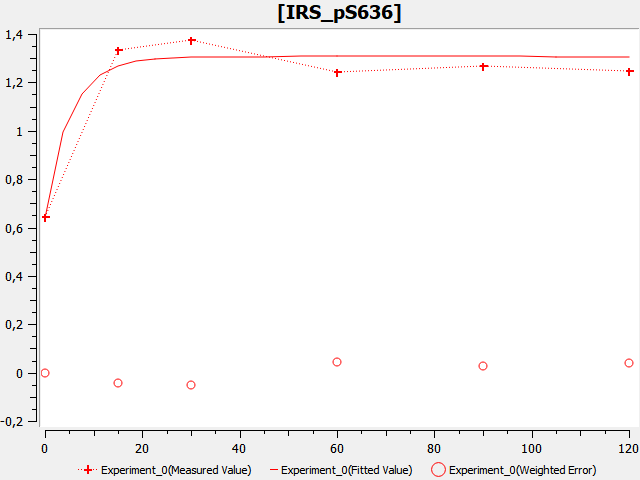
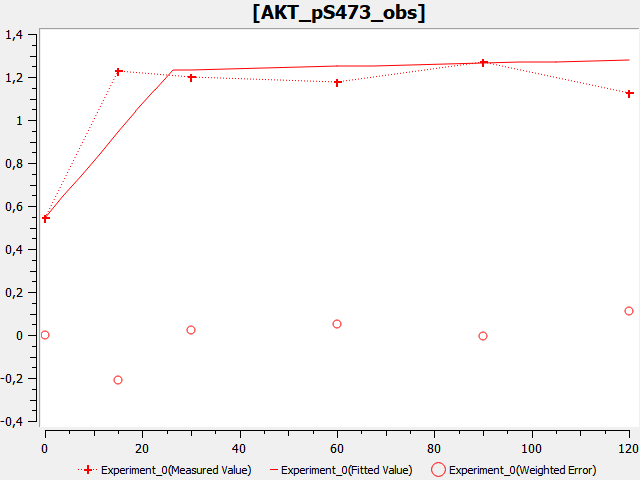
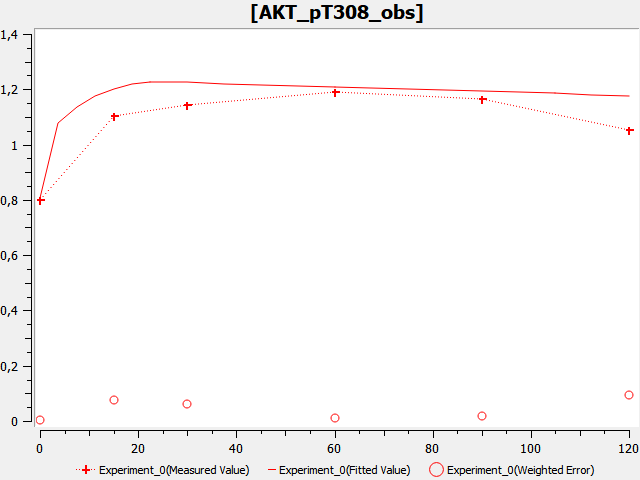
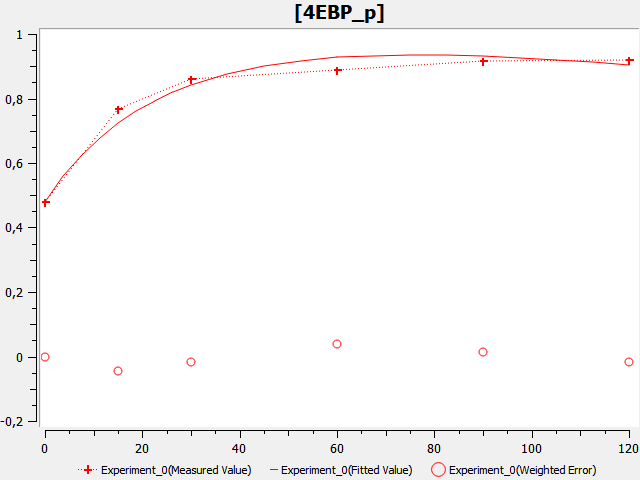


Figure 3. Model estimates using Michealis-Menten kinetics

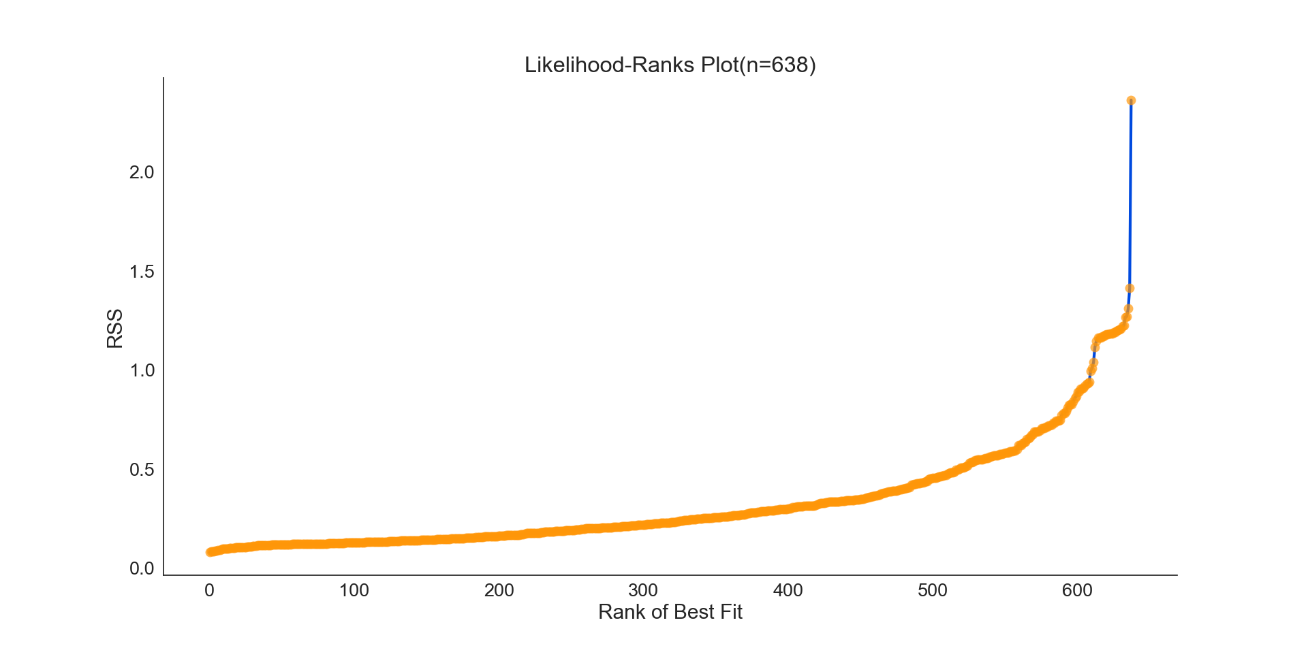


Figure 4. Log-likelihood-Ranks plot for the model topology shown in figure 1 using a GA

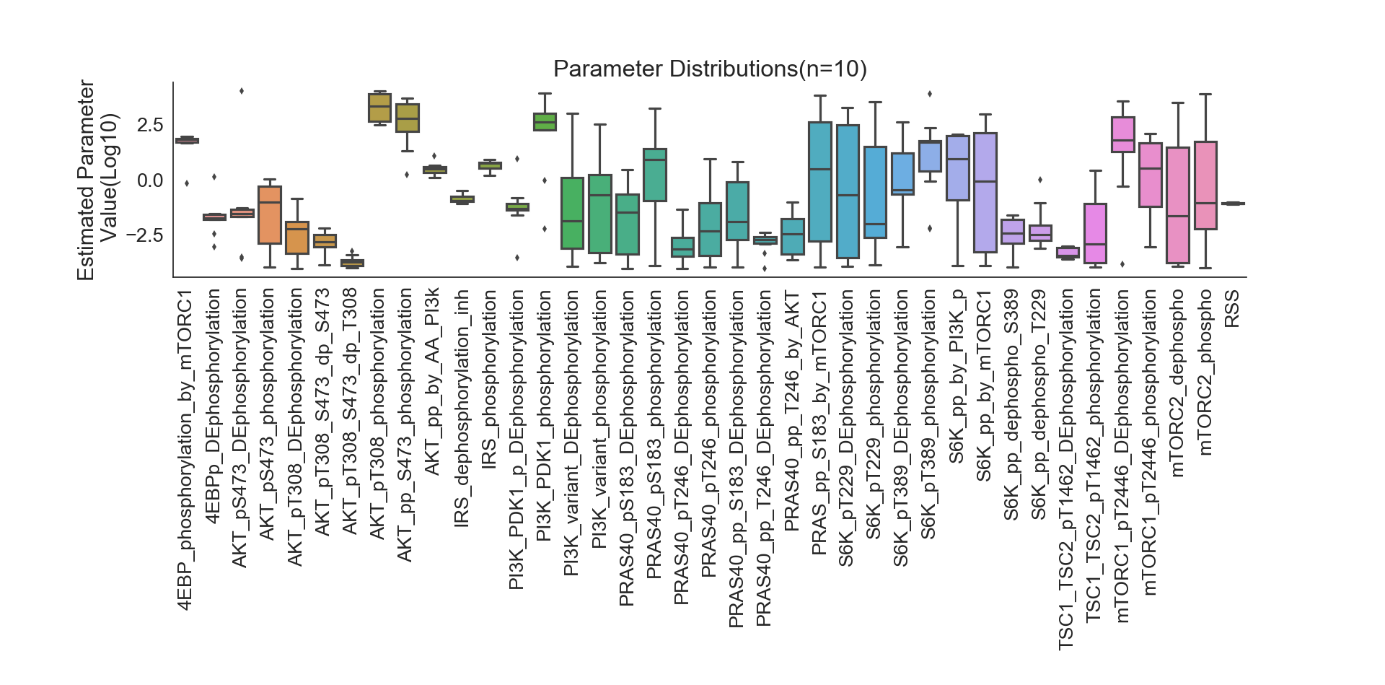
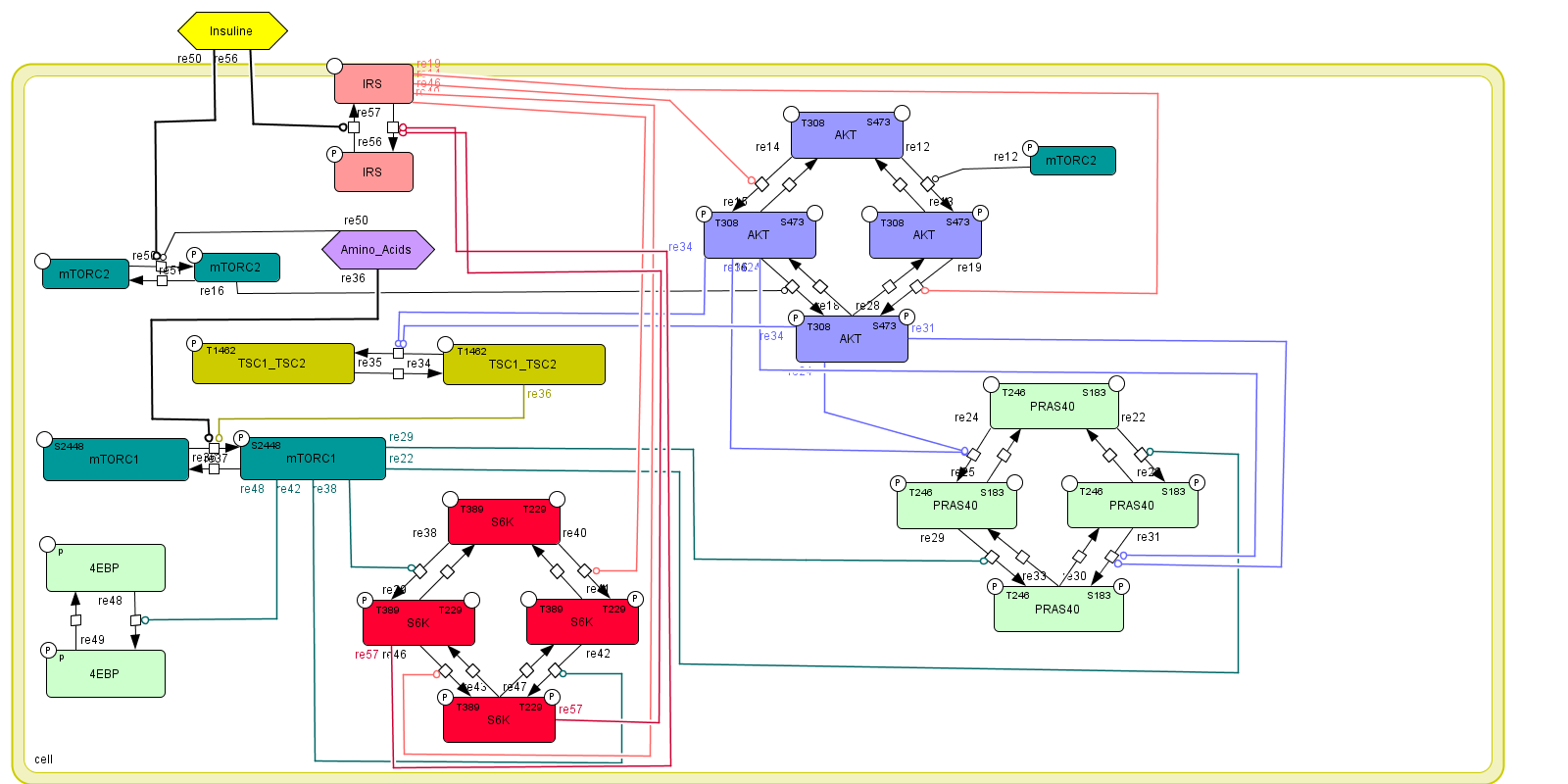
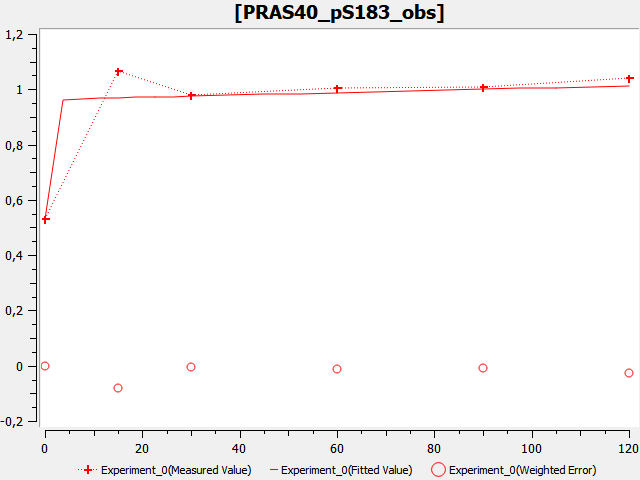
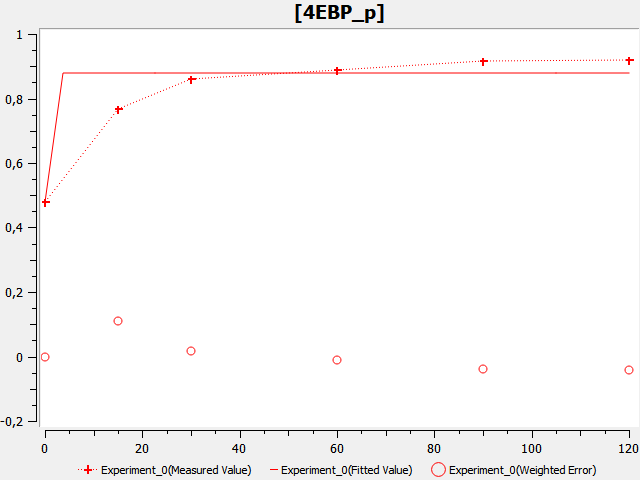
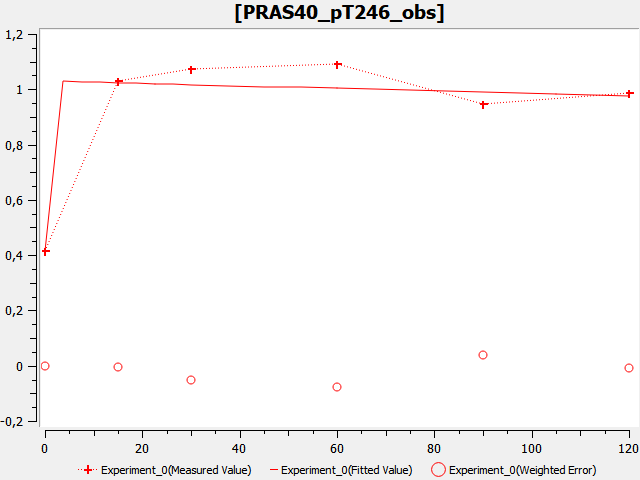
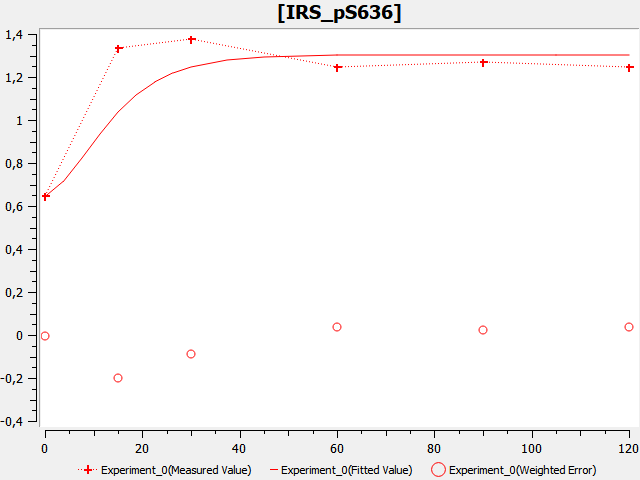
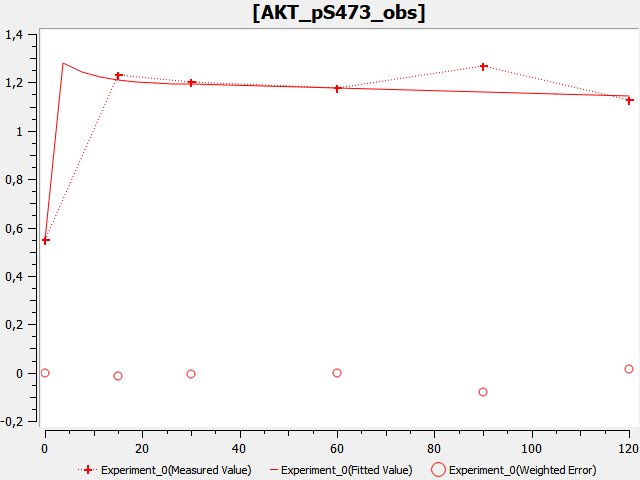
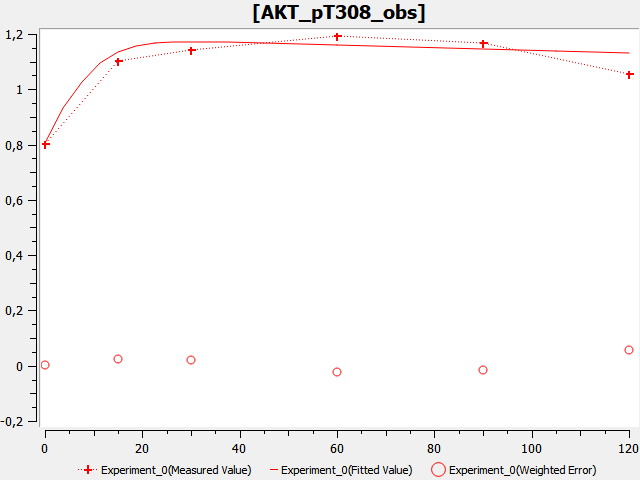


Figure 6. Boxplot showing distribution of parameters from the parameter estimation.

In addition to running the model shown in figure 1 with Michealis-Menten kinetics another model with the PI3K reactions were removed. This was done as they have larger parameter distributions indicating a low chance for identifiability. In addition a model with fewer parameters that performs equally well increases the chance of identifiability for the other parameters. The model topology is shown in figure 7. Parameter estimations were performed using a GE with 500 generations with a population size of 75. Time course resulting from the estimations seemed to be able to capture the trends as shown in figure 8. Note that one could also try and only remove the PI3K\_variant as the PI3K seems to have a narrow distribution.



Figuur 7. Reduced model topology from figure 1.

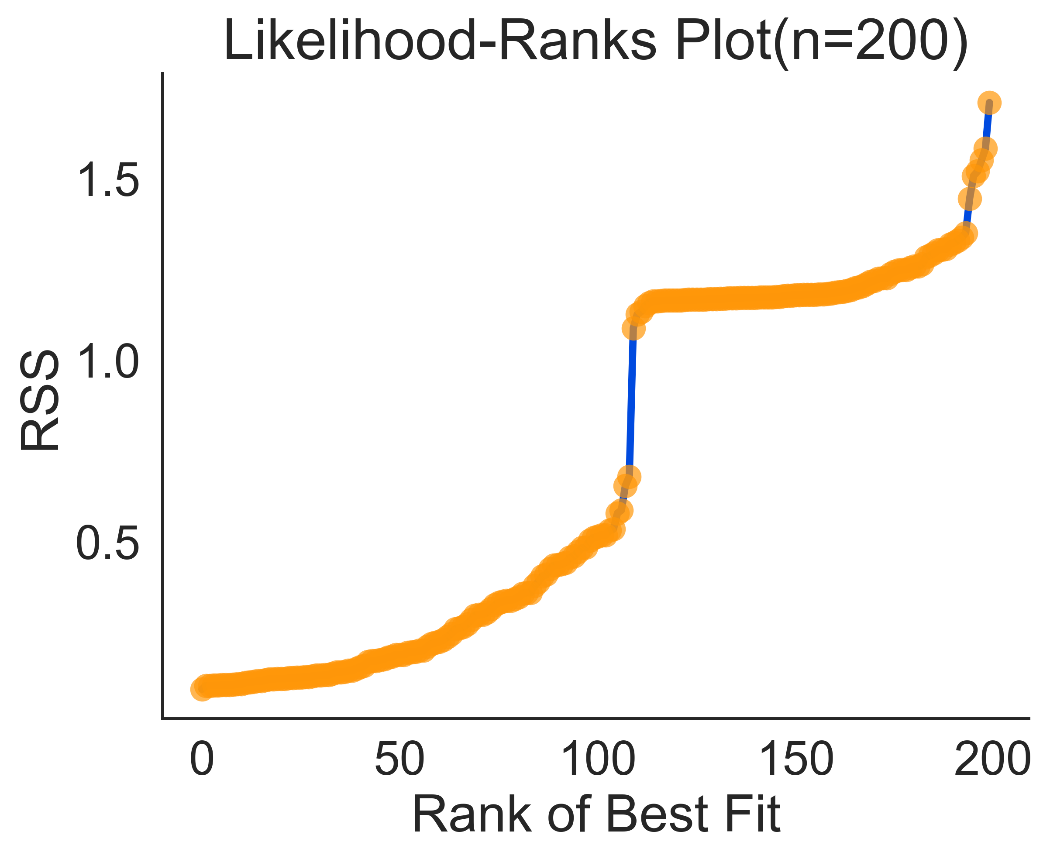


Figuur 8. Some predictions of the less complex model topology shown in figure 7.

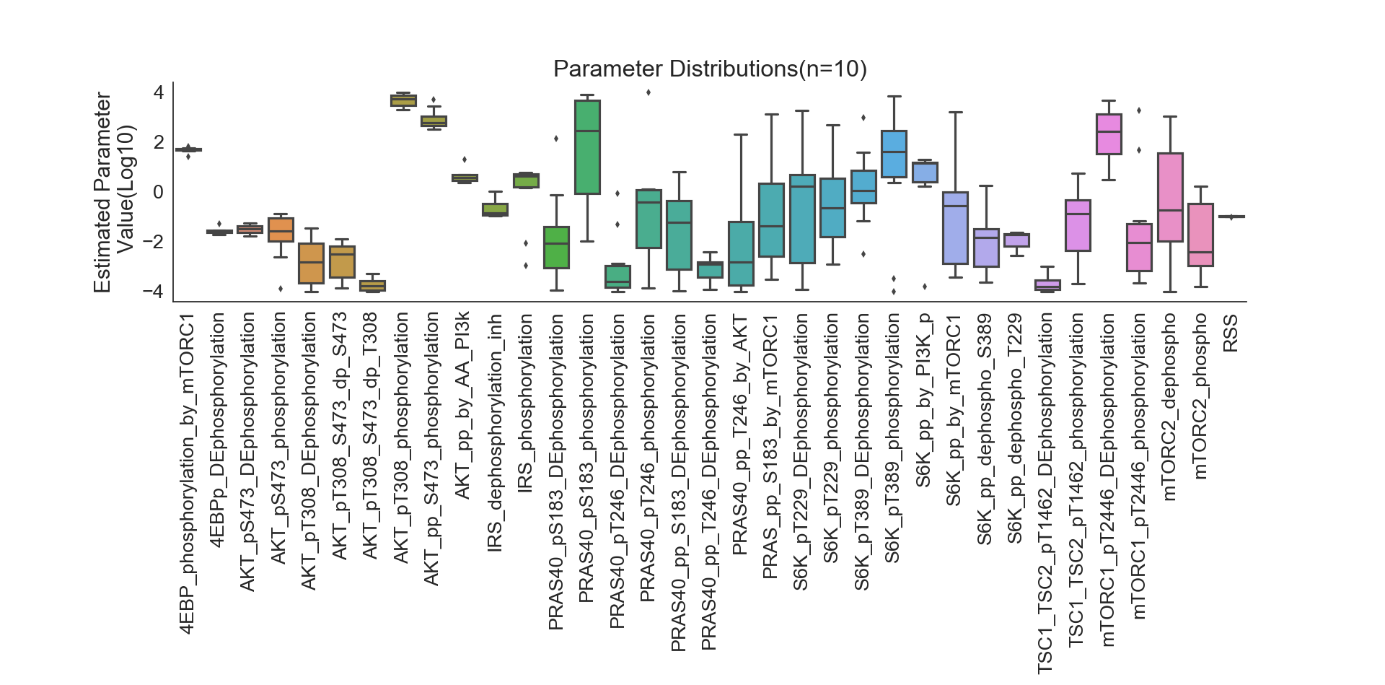
For checking if the new model topology also works after multiple parameter estimation 200 GE with 500 generations with a population of 75 where run. The resulting log-likelihood-ranks plot is shown in figure 9 and the resulting parameter distributions are seen in figure 10. In these figures it can be seen that the global search could be improved by increasing the number of generations and population sizes or one could do a particle swarm to get closer to the global optima. In addition, the boxplot show similar distributions for the estimated parameters. Judging from the boxplots the removal of the PI3K’s seem to have little effect on the model estimation. Hence it could be a potential model for a more formal comparison one should do a model comparison and calculate the Bayesian information criterion (BIC) and Akaine information criterion (AIC). The expectation is that the second model will have better values as it has similar fits with fewer parameters.

Up until the last two models parameter distribution where wide and the identifiability analysis that where run gave practically unidentifiable results. The current two topologies seem promising and at the moment identifiability analysis are being run in parallel to a model comparison. These analysis will provide more concrete insights in the performance of the models. If the better of the two models is identifiable one can undertake the next step to change initial concentrations and look if the same model can be used for different cell lines.

In the first model there where two identifiable parameters and many close to or could be argued as identifiable. An overview can be seen in appendix 2. Based on the output of the identifiability analysis I would advise to fix the parameters that have been shown to be identifiable and repeat the multiple parameter estimation using a stronger algorithm such as the particle swarm using a swarm size of 50 and an iteration limit of 2000 followed by a HJ with an iteration limit of 1000 and tolerance of 1e-10.



Figuur 9. Log likelihood-ranks plot of the reduced model topology using GA.



Figuur 10. Parameter distributions of the reduced model topology.

## Appendix 1

Tabel 1

|  |  |  |
| --- | --- | --- |
| **Model number** | **Description** | **Why change** |
| **1** | Initial model. | No double phosphorylation, topology mistakes with S6K feedback and phosphorylation |
| **2** | Initial model based on rate questions without global quantities | Input of TSC2 required to keep model running. Amino acid depletion. |
| **3** | Model 1 with added insulin input. Different mtorc1/2 states added to cope with TSC2 change. Amino acid input added | Predictions work however due to increasing insulin input the model predictions deteriorate later time points |
| **4** | Insulin/amino acid output removed, mTORC1 state added | Non-identifiable results. |
| **5** | Model changed towards one with global quantities rather than specie based rate equations. This overcame the amino acid/insulin input/output | mTOR activated states are very complex, introducing many parameters.  No modifier rate equations |
| **6** | Model adapted to incorporate modifier effects in rate equations. IRS\_PDK1 introduced | Model was becoming over complex for the mTORC1/2  Non-identifiable results. |
| **7** | mTROC1/2 simplified | topology issues regarding PI3K, S6K and IRS |
| **8** | Topology issues resolved | Potential identifiability issues due to the many nodes of which no experimental data is available. |
| **9** | PI3K\_variant and PI3K\_PDK1 removed from model to reduce complexity. |  |

Table 2. Overview of identifiability parameters from the model topology shown in figure 1.

|  |  |  |
| --- | --- | --- |
| **Parameter** | **Identifiable** | **Non-identifiable** |
| **4EBP\_phosphorylation\_by\_mTORC1** | Almost |  |
| **4EBPp\_DEphosphorylation** | Almost |  |
| **AKT\_pp\_by\_AA\_PI3k** | X |  |
| **AKT\_pp\_S473\_phosphorylation** |  | X |
| **AKT\_pS473\_DEphosphorylation** |  | X |
| **AKT\_pS473\_phosphorylation** |  | X |
| **AKT\_pT308\_DEphosphorylation** |  | X |
| **AKT\_pT308\_phosphorylation** |  | X |
| **AKT\_pT308\_S473\_dp\_S473** |  | X |
| **AKT\_pT308\_S473\_dp\_T308** |  | X |
| **IRS\_dephosphorylation\_inh** | X |  |
| **IRS\_phosphorylation** | Arguable |  |
| **mTORC1\_pT2446\_DEphosphorylation** |  | X |
| **mTORC1\_pT2446\_phosphorylation** |  | X |
| **mTORC2\_dephospho** |  | X |
| **mTORC2\_phospho** |  | X |
| **PI3K\_PDK1\_p\_DEphosphorylation** | Almost |  |
| **PI3K\_PDK1\_phosphorylation** |  | X |
| **PI3K\_variant\_DEphosphorylation** |  | X |
| **PI3K\_variant\_phosphorylation** |  | X |
| **PRAS\_pp\_S183\_by\_mTORC1** |  | X |
| **PRAS40\_pp\_S183\_DEphosphorylation** |  | X |
| **PRAS40\_pp\_T246\_by\_AKT** |  | X |
| **PRAS40\_pp\_T246\_DEphosphorylation** |  | X |
| **PRAS40\_pS183\_DEphosphorylation** |  | X |
| **PRAS40\_pS183\_phosphorylation** |  | X |
| **PRAS40\_pT246\_DEphosphorylation** |  | X |
| **PRAS40\_pT246\_phosphorylation** |  | X |
| **S6K\_pp\_by\_mTORC1** |  | X |
| **S6K\_pp\_by\_PI3K\_p** | X |  |
| **S6K\_pp\_dephospho\_S389** |  | X |
| **S6K\_pp\_dephospho\_T229** | almost |  |
| **S6K\_pT229\_DEphosphorylation** |  | X |
| **S6K\_pT229\_phosphorylation** |  | X |
| **S6K\_pT389\_DEphosphorylation** |  | X |
| **S6K\_pT389\_phosphorylation** |  | X |
| **TSC1\_TSC2\_pT1462\_DEphosphorylation** |  | X |
| **TSC1\_TSC2\_pT1462\_phosphorylation** |  | X |