**Assessment of Model Sensitivity**

**Manipulation of Optimization Parameters**

GRNmap uses the optimization parameters kk\_max, MaxIter, TolFun, MaxFunEval, and TolX to optimize the fit of the model. The parameter kk\_max is the number of times that the optimization loop is rerun, which can improve the model performance. The MaxIter is the maximum iterations that the model will run. This parameter should be high enough that the model does not stop before completing the estimation. Therefore, if the MaxIter is too low, the model may not complete the optimization. The TolFun is the difference between the least squares of each rerun that signifies when the model is no longer making improvements. The MaxFunEval is the maximum number of times that the function will evaluate the least squares cost. Lastly, the TolX indicates the maximum difference between the least errors cost before the program signifies there is no more improvement of the model (Dahlquist et al., 2018).

The standard optimization parameter values used in the model are 1 for kk\_max, 1.00e8 for MaxIter, 1.00e-6 for TolFun, 1.00e8 for MaxFunEval, and 1.00e-6 for TolX. In order to assess the sensitivity of the model to changes in these optimization parameters, each parameter was changed by an order of magnitude in either direction and then run. The LSE:minLSE ratios of the models were compared to that of the standard model run with the midpoint values. Deviation from the standard model indicates that either the model ran better, as in the case of a decrease in the LSE:minLSE ratio, or worse, indicated by an increase in the ratio. In addition, the iteration counts of the models were compared to determine whether the changes to the optimization parameters affected the number of times the model evaluated the least squares before stopping the model run.

***Methods***

Using db5 all-strain data, the optimization parameters in the input Excel workbook were changed by one degree of magnitude in each direction for kk\_max, MaxIter, TolFun, MaxFunEval, and TolX (For example, the MaxIter standard value is 1.00e8 and two models were run where MaxIter was set to 1.00e9 and one set to 1.00e7). One optimization parameter was changed per new model, with the other parameters set with their standard values.

The resulting ten new model input files were run through GRNmap and the LSE:minLSE ratios and iteration counts of each model were then compared to determine if the changes to the optimization parameters affected the fit of the model. The standard model, where none of the optimization parameters were changed, was the db5 all-strain model run. The model run where kk\_max was set to 0.1 failed to complete the estimation, therefore, it was not included in the comparison.

***Results/ Discussion***

The LSE:minLSE ratio for the run under standard conditions (mid-point value for each parameter) was 1.4081816. When the parameters were changed by one magnitude in either direction, the ratio changed only for kk\_max at 10 and TolX at 1.00E-07, but the change was minimal with a 2.56e-4% decrease for kk\_max at 10 and a 3.18e-4% decrease for TolX at 1.00E-07. The other runs caused no difference in the ratio. Therefore, changing the value for the optimization parameters made very minimal difference to the fit of the model.

The iteration count of the model indicates the number of times the least squares is evaluated by the program before the model run stops (Dahlquist et al., 2018). The iteration count for the run under standard conditions (mid-point value for each parameter) is 111,242. Again, changing the kk\_max to 10 and the TolX to 1.00e-7 caused a change in the iteration count. The change in iteration count for these parameters was greater than the change that occurred for the LSE:minLSE ratio, especially for kk\_max which showed a 480% increase, while the change in TolX caused a 3.5% increase (Fig 1). This increase could be expected however, because changing kk\_max to 10 increases the number of times the optimization loop is rerun, which in turn can increase the number of iterations the model runs through. Likewise, decreasing the TolX by one magnitude decreases the maximum allowed difference in the least squares cost before the model stops, which may cause the model to run through more iterations.

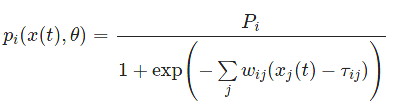
Overall, these results indicate that changing the optimization parameters does not have a major impact on the fit of the model, while the iteration count is greatly affected by increasing the kk\_max value.

|  |
| --- |
| **Fig 1:** The LSE:min LSE ratio remained the same for all of the optimization parameter tests except for when kk\_Max was set to 10 and TolX was set to 1.00e-7 in which case, the ratio decreased. Likewise, the iteration count remained the same except for these same optimization parameter changes. The iteration count increased in both of these cases. The mid-value model is the one run with the standard values for each optimization parameter. |

**Variably Fixed or Estimated Weights, Thresholds, and Production Rates**

GRNmap performs estimations of regulatory weight, expression threshold, and production rate parameters. In a standard model, each of these parameters is estimated through the program, but the threshold and production rate parameters can be input as fixed values. When fixed, the initial guesses for the production rate and/or threshold will remain throughout the model estimation, which could affect the fit of the model. Therefore, in order to determine the impact of fixing or estimating these parameters, models were run where production rate and threshold were either estimated or fixed.

When estimated, these parameters are optimized using ordinary differential equations in GRNmap. The production rate is estimated using the sigmoidal function:

, where *Pi* is the initial guess production rates and *pi* is the estimated rate (Dahlquist et al., 2015). The expression over time is then modeled by production minus degradation rate. The initial guess for the production rate was derived through doubling the degradation rates from the Neymotin et al. (2014) paper.

The threshold indicates the point at which activators either induce production, when expression is above threshold, or stop production, when expression is below threshold. For repressors, production is induced when the expression is below the threshold level, while production is stopped when the expression is above threshold (Dahlquist et al., 2015). In the model, threshold is estimated using:

, where *wi j* is the weight of a regulatory relationship and τ*i j* is the threshold at which the production turns off or on (Dahlquist et al., 2015). In standard model runs, the initial guess for the threshold was set to zero.

***Methods***

Using db1-db7 all-strain data input files, six new modified models per network were created where thethresholds *(b)* and production rates *(P)* were either estimated or fixed in the model and the weights *(w)* were estimated in each new model (Table 1). For the models where the production rates were fixed, additional models were run where the production rates used as initial guesses were derived from the Neymotin et al. (2014) paper. For the models that were run with fixed threshold values, the initial guess was set to 0. The models were run in GRNmap and from the output, the LSE:minLSE ratio was calculated for each model in order to assess and compare the fit of the models.

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 1:** Fix/estimate parameter model runs with indication of which parameters were fixed or estimated. Weights were estimated for all model runs, while the threshold and production rate were variably estimated or fixed between the runs. The initial guess production rates were derived by doubling the Neymotin et al. (2014) degradation rates. Neymotin production rates were derived from the Neymotin et al. (2014) paper and used in the models where indicated. The threshold initial guess was set to 0. | | | |
| **Model Run** | **Weight *(w)*** | **Threshold *(b)*** | **Production Rate *(P)*** |
| **Est *P, b, w*** | Estimated | Estimated | Estimated |
| **Est *P, w;* Fix *b*** | Estimated | Fixed | Estimated |
| **Est *b, w;* Fix *P*** | Estimated | Estimated | Fixed |
| **Est *b, w;* Fix *P*** (using Neymotin Production Rates) | Estimated | Estimated | Fixed |
| **Est *w;* Fix *P, b*** | Estimated | Fixed | Fixed |
| **Est *w;* Fix *P, b*** (using Neymotin Production Rates) | Estimated | Fixed | Fixed |

***Results/ Discussion***

In order to determine the effect of estimating or fixing parameters on the fit of the model, the LSE:minLSE ratios from each model run were compared. The lower the LSE:minLSE ratio, the better the fit of the model. Db5 was chosen as the candidate network for analyzing the effect on model sensitivity, as db1-db7 model runs resulted in the same trends for increased or decreased LSE:minLSE ratio among the estimate/fix model runs.

The lowest ratio occurred for the models where *w, b,* and *P* were all estimated, therefore, it was determined that fixing any parameter decreased the fit of the model (Fig 1). Estimating or fixing the threshold did not change the LSE:minLSE ratio value greatly, which indicates that threshold does not have a major impact on the fit of the model. Conversely, fixing the production rates made the LSE:minLSE ratio increase more, indicating their importance on the fit of the model (Fig 1). In addition, the use of the Neymotin et al. (2014) production rates as the fixed values caused the LSE:minLSE ratio to increase even more, indicating that the initial guesses for the production rates are a better fit for the model than the Neymotin et al. (2014) values.

|  |
| --- |
| **Fig 2:** The candidate db5 network LSE:minLSE ratios for the estimate/fix production rate (*P*), weight (*w)*, and threshold (*b)* model runs. Production rates used were the initial guess rate or the Neymotin et al. (2014) production rate (designated by “Neymotin”). For db1-db7, fixing production rate caused a greater ratio, the greatest of which were fixing the Neymotin production rates. |

**Conclusion**

From these sensitivity assessments of the model, it can be determined that changes in the optimization parameters of the model do not impact the fit of the model, but an increase in the kk\_max value does increase the iteration count. Similarly, it was determined that although estimating the weights, thresholds, and production rates resulted in the best fit of the model, fixing the threshold parameter did not impact the model dramatically. Fixing the production rate as the initial guesses caused the model to perform worse, but the worst fit came from the models that were run with the Neymotin et al. (2014) production rates. This indicates that our initial guess production rates are a better fit for the model than the Neymotin values. Overall, it was determined that the model is not too sensitive to changes in the optimization parameter, but is sensitive to fixing the production rates

**References:**

Dahlquist, K.D., Fitzpatrick, B.G., Camacho, E.T., Entzminger, S.D., & Wanner, N.C. (2015). Parameter estimation for gene regulatory networks from microarray data: cold shock response in *Saccharomyces cerevisiae. Bulletin of Mathematical Biology,* 77(8), 1457-1492. doi:10.1007/s11538-015-0092-6

Dahlquist, K., Fitzpatrick, B., & Dionisio, J. D. (2018, August). GRNmap (Gene Regulatory Network modeling and parameter estimation). Retrieved February 13, 2020, from https://kdahlquist.github.io/GRNmap/documentation/

Neymotin, B., Athanasiadou,R., & Gresham, D. (2014). Determination of in vivo RNA kinetics using RATE-seq. *RNA*, 20(10), 1645-1652. doi:10.1261/rna

**Appendix:**

[**https://docs.google.com/document/d/1wLl84bKPZzHhDU8if-JbpDY-ZwwNRsNavYkZaoVv8xw/edit**](https://docs.google.com/document/d/1wLl84bKPZzHhDU8if-JbpDY-ZwwNRsNavYkZaoVv8xw/edit)