Test-Driven Development and Functionality Improvements to GRNmap, a Gene Regulatory Network Modeling Application

LMU LA McNair Scholars Program

Trixie Anne Roque[1], Kam D. Dahlquist[2], Ben G. Fitzpatrick[3] , and Dr. John David N. Dionisio[1]

[1]Department of Electrical Engineering and Computer Science, [2]Department of Biology, [3]Department of Mathematics,
Loyola Marymount University, 1 LMU Drive, Los Angeles, CA 90045 USA

GRNmap Uses ODE to Model Networks

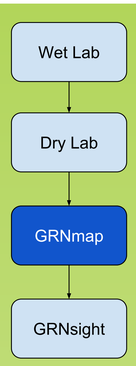


Figure 1: Laboratory data flow

$$\frac{dx_i(t)}{dt} = \frac{P_i}{1+\exp(-\sum_j(w_{ij}x_j(t))+b_i)} - d_i x_i(t)$$

Figure 2: Sigmoidal model used by GRNmap

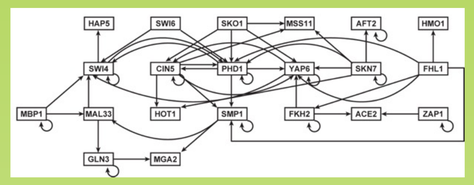


Figure 3: Example of a gene regulatory network

GRNmap is a Matlab software that models gene regulatory network dynamics. It accepts laboratory microarray data that is pipelined and processed by our various groups as shown in Figure 1:

- The wet lab generates gene expression data using DNA Microarrays.
- The dry lab works on statistical analysis of the data and generates gene regulatory networks (GRN). The data is presented as an Excel workbook.
- The GRN is run through GRNmap which estimates production rates, weights, and thresholds using 1 of 2 ordinary differential equations (one is shown in Figure 2).
- The data outputted by GRNmap is used by GRNsight to visualize the network. A visual depiction of a GRN is shown in Figure 3:
  - Each node represents a gene
  - Each edge represents a regulatory relationship between the genes
  - The network is represented as an adjacency matrix

GRNmap Functionality Changes

Over the course of its development, we have improved on the functionalities of the software package. Since v1.0.8, changes we have made to the source code include:

- New Features:
  - Changed the names of the worksheets in the input and output workbooks
  - Computed standard deviations
  - Optimization diagnostics outputted for each run (contains LSE, penalty term, and iteration count)
  - Computed minimum LSE and sum of squares error of individual genes
  - Plots are saved according to their gene names
- Bugs fixes:
  - Correctly outputting estimated production rates and threshold parameters worksheets
  - Corrected computation for threshold for genes with no inputs
  - Corrected penalty computation for production rates
- Test-driven development:
  - The standard process is shown in Figure 4
  - Created 16 manual input test sheets to compare outputs with when running the code. A visual representation of all 16 is shown in Figure 5
  - Created some minor unit tests for checking if correct outputs are present. Pseudocode for what the unit tests look like is shown in Figure 6

These changes have been documented in our external GRNmap website (http://kdahlquist.github.io/GRNmap/index.html) and GRNmap developer wiki (https://github.com/kdahlquist/GRNmap/wiki)

Test-Driven Development Process

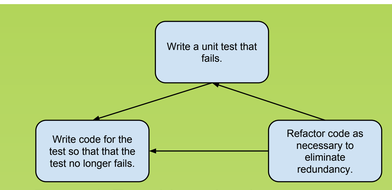


Figure 4: Test-driven development process

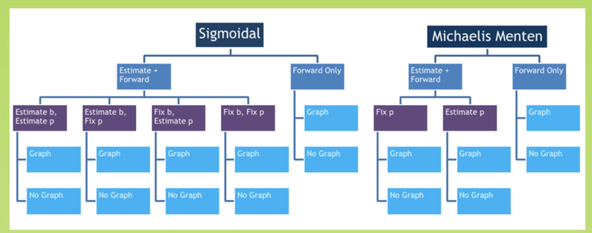


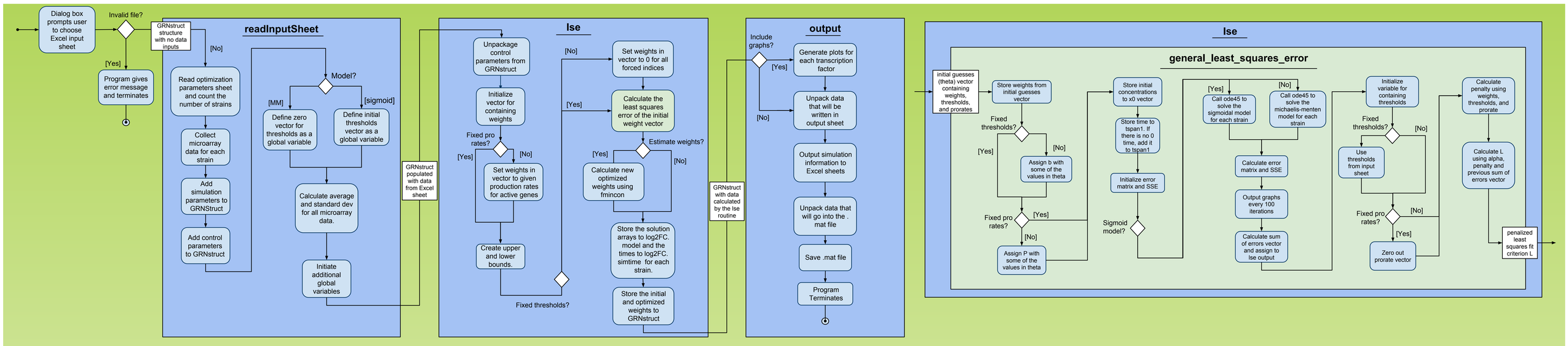
Figure 5: 16 manual test input sheets

```
DEFINE main function
    CALL functiontests (localfunctions) to make a tests array
END

DEFINE function firstTest (testCase)
    actualOutput = evaluate function by using known inputs
    expectedOutput = assign expected results
    VERIFY actualOutput equals expectedOutput
END
```

Figure 6: Pseudocode for unit tests

Activity Diagram Shows How Data is Processed by GRNmap



References

- Dahlquist, K.D., Fitzpatrick, B.G., Camacho, E.T., Entzminger, S.D., and Wanner, N.C. (2015) Parameter Estimation for Gene Regulatory Networks from Microarray Data: Cold Shock Response in Saccharomyces cerevisiae. Bulletin of Mathematical Biology, in press.

Future Work

- Complete the testing framework for all current functionality of the code, fixing bugs and refactoring code as needed
- Clean up the variable names, and how the program discovers which strains are present and which genes are deleted
- Move the documentation for how to use GRNmap from the wiki to the documentation page of the web site
- Add functionality so that GRNmap computers the within- and between-strain ANOVA values

Acknowledgments

We would like to thank Juan Carrillo, Nicholas A. Rohacz, Alondra Vega, Stephanie D. Kuelbs, Nathan C. Wanner, and Erika T. Camacho for previous work on the GRNmap program. This project was supported by the Summer Undergraduate Research Program at Loyola Marymount University (J.S.C.), NSF-DMS award #0921038 (K.D.D., B.G.F., and K.S), and the Clarence Wallen, S.J. Chair in Mathematics (B.G.F.).