Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of *Saccharomyces cerevisiae*

Vu, T. T., & Vohradsky, J. (2007). Nonlinear differential equation model for quantification of transcriptional regulation applied to microarray data of Saccharomyces cerevisiae. *Nucleic acids research, 35*(1), 279-287. doi: 10.1093/nar/gkl1001

Conor Keith and Nika Vafadari

Department of Biology
Loyola Marymount University

BIOL 398-05/MATH 388-01
March 30, 2017

Outline

- Goal: Identification of transcriptional regulators that control cell cycle regulated genes of Saccharomyces cerevisiae using non linear model
- A new nonlinear differential equation model of gene expression
- Dynamic transcriptional control and optimization using least squares minimization
- Algorithm utilizes dynamic model of time continuous gene expression
- Method for determining correct regulators
- Trends in expression profiles for selected genes and their regulators
- Compared to linear model, nonlinear model gives better results in terms of correct identification of the regulator and better fit to gene expression profile of target.

Transcriptional regulation leads to changes in gene expression

- **Transcriptional regulatory proteins recognize specific promoter sequences and allow binding of RNA polymerase and initiating transcription**
- **DNA microarrays allow for the visualization and recording of changes in gene expression over time**
- **Changes in gene expression throughout the cell cycle of *Saccharomyces cerevisiae* provides insight into regulator-target gene relationships and the network interactions that result**
- **Therefore, many studies focus on analyzing microarray data through clustering methods to identify cell cycle controlled genes**

Previous Methods Identify Upstream Regulatory Genes through Generalized Linear Model

- **Wolf and Wang: used fuzzy logic**
- **Nachman *et al.*: used dynamic Bayesian networks with a kinematic model**
- **Bar-Joseph: used gene expression analysis and genomic info alongside one another**
- **Wang *et al.* and Makita et al.: extending the work of Bar-Joseph, incorporated promoter sequence analysis into gene expression analysis**

Alternative Method Replaces Linear Model with Non Linear Model

- 184 potential regulators chosen
- Set of 40 specific target genes within *S. cerevisiae* selected
- Genes from within set of potential regulators selected and applied to model in order to see if regulator fits target gene expression profile correctly
- Repeated for other target gene and potential regulator combos
- True regulators determined by identifying regulators that model the target gene profile correctly and are found in YEASTRACT database

Outline

- Goal: Identification of transcriptional regulators that control cell cycle regulated genes of Saccharomyces cerevisiae using non linear model
- Dynamic transcriptional control and optimization using least squares minimization
- Algorithm utilizes dynamic model of time continuous gene expression
- Method for determining correct regulators
- Trends in expression profiles for selected genes and their regulators
- Compared to linear model, nonlinear model gives better results in terms of correct identification of the regulator and better fit to gene expression profile of target.

Dynamic Model for Transcriptional Control

- g: regulatory effect for a particular gene
- w: regulatory weights
- y: expression levels of regulators
- b: transcription initiation delay
- ϱ: regulatory effects of non-target genes
- k: rate constants
- z: target gene expression
- j: regulators

$$g \approx \sum_j w_j y_j - b, \qquad \mathbf{1}$$

$$\rho = \frac{1}{1 + \exp\left(-\sum_{j=1..m} w_j y_j + b\right)}, \qquad \mathbf{2}$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = k_1 \frac{1}{1 + \exp\left(-\sum_{j=1..m} w_j y_j + b\right)} - k_2 z. \qquad \mathbf{3}$$

Simplified Model for One Transcriptional Control

- Simplified version of equation three, focusing on the case of only one transcriptional factor
- Polynomial coefficients [a_0,...,a_n] are computed from the gene expression profile using least square minimization.
- Polynomial used as an approximation for 'true' expression profile with experimental errors.

$$\frac{dz}{dt} = \frac{k_1}{1 + \exp(-wy + b)} - k_2 z, \quad \mathbf{4}$$

$$y \approx a_0 + a_1 t + a_2 t^2 + \ldots + a_n t^n. \quad \mathbf{5}$$

Mean Square Error Function

- $\mathbf{Z}\ \{z(t_\tau)\}$: Expression profiles of the target genes
- $\mathbf{Y}\{y(t_\tau)\}$: Expression profiles of the regulator genes
- Time points $t = \tau$, $\tau = 1,2,\ldots, Q$
- Search for gene profiles: $\mathbf{Y} \in \{\mathbf{Y}_i, 1,2,\ldots,m\}$ (the m pool of regulators) that minimize E

$$E = \frac{1}{Q} \sum_{\tau=1}^{Q} \left[z(t_\tau) - z^c(t_\tau)\right]^2, \qquad 6$$

$$\frac{\mathrm{d}z}{\mathrm{d}t} = \mathrm{d}_0 + \mathrm{d}_1 y - \mathrm{d}_2 z, \qquad 7$$

Outline

- Goal: Identification of transcriptional regulators that control cell cycle regulated genes of Saccharomyces cerevisiae using non linear model
- Dynamic transcriptional control and optimization using least squares minimization
- Algorithm utilizes dynamic model of time continuous gene expression
- Method for determining correct regulators
- Trends in expression profiles for selected genes and their regulators
- Compared to linear model, nonlinear model gives better results in terms of correct identification of the regulator and better fit to gene expression profile of target.

Computational Algorithm

- **Goal: select a set of potential regulators of a particular target gene by estimating its expression profile**
- **Method: find possible regulators using least squares minimization and the model equation 4 to minimize error function**
- **Degree of polynomial (equation 5) is chosen according to number of data points in in the profile and the level of fluctuations for each experiment.**
- **Differential equation 4 is solved numerically and parameters are optimized in least squares loop until desired predifined number of iterations is obtained.**

Computational Algorithm Continued

1. Use equation 5 to fit regulator genes with polynomial of degree n
2. Choose specific target gene
3. Choose possible regulator from large pool of potential ones
4. Use least squares minimization (equation 4) on target gene/ regulators & error function (equation 6)
5. Repeat starting at step 3 for all potential regulators
6. Choose best fit regulators that match criteria
7. Repeat for all the target genes starting at step 2

Dataset Selected Based on Previous Work

- Used eukaryotic cell cycle dataset (Spellman et al.) to evaluate model which includes..
  - gene expression changes at 18 time points over 2 cell cycle periods
  - 6178 open reading frames on microarray chip
  - identified 800 genes associated with cell cycle according to their expression
- However, number of regulators controlling cell cycle < 800, therefore pool of 184 chosen by researchers for this experiment based on YEASTRACT database and previous papers
- Chose same 40 target genes from Chen et al. paper to compare data

Inference of Regulators

- Data put into log base 2 of ratio [actual value of mRNA divided by value of a standard]
- Prior to use of algorithm, data squared and least squares minimization applied to target gene for potential regulator

$$z^p \approx c_0 + c_1 t + c_2 t^2 + \ldots + c_n t^n, \tag{8}$$

- z^p = approximation of the unknown real profile of target gene
- Approximation takes into account error due to experimentation and natural fluctuations by polynomial fit

Calculation of Deviation from Experimental Data Allows for Identification of Best Regulators

- In order to identify most probable regulator for target gene, must determine which regulator profile best models target gene profile (Equation 4) and minimizes error (Equation 6)
- Therefore, chose regulators with an E less than or equal to the deviation E1
- Best regulators= those that have a recognizably smaller E than others

$$E_1 = \frac{1}{Q}\sum_{\tau=1}^{Q}\left[z(t_\tau)-z^p(t_\tau)\right]^2. \tag{9}$$

Outline

- Goal: Identification of transcriptional regulators that control cell cycle regulated genes of Saccharomyces cerevisiae using non linear model
- Dynamic transcriptional control and optimization using least squares minimization
- Algorithm utilizes dynamic model of time continuous gene expression
- Method for determining correct regulators
- Trends in expression profiles for selected genes and their regulators
- Compared to linear model, nonlinear model gives better results in terms of correct identification of the regulator and better fit to gene expression profile of target.

Comparison to YEASTRACT Database Determines Correct Regulators

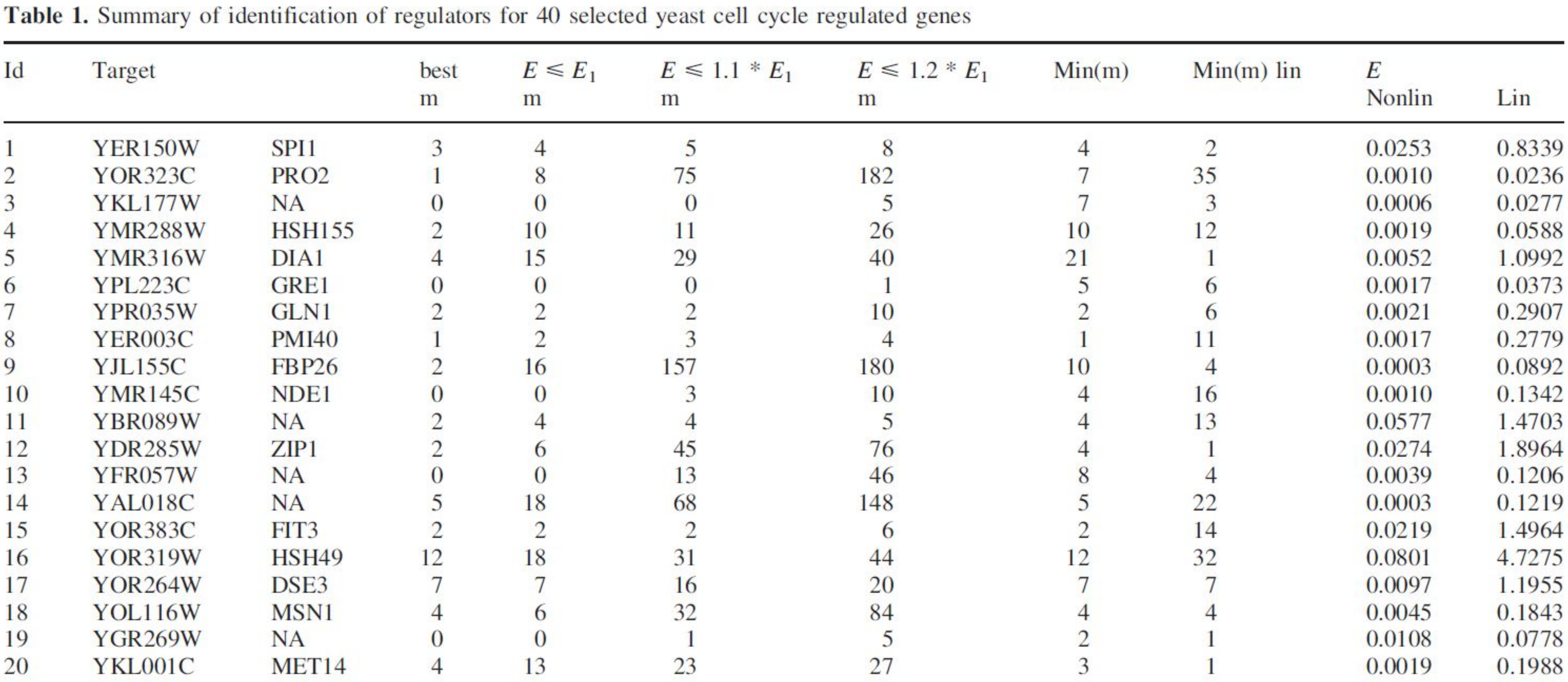
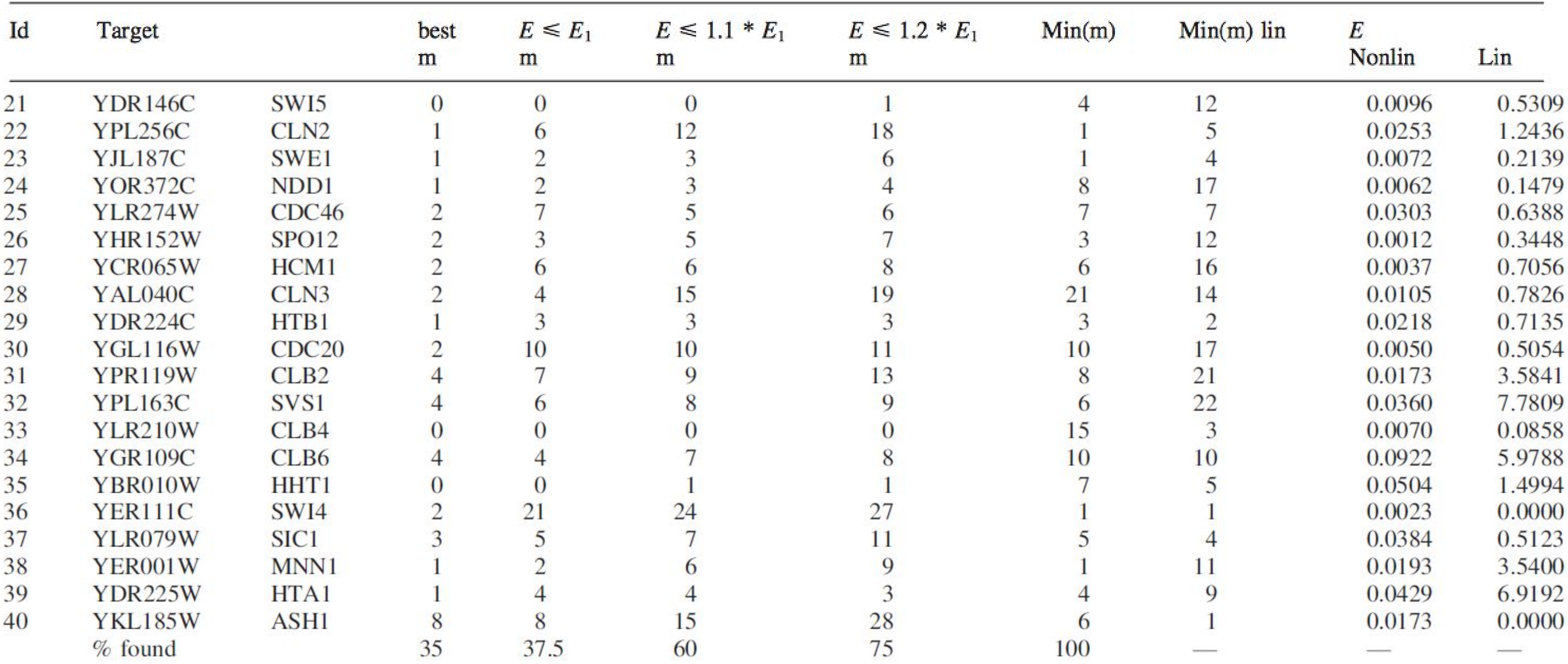


Table 1. Summary of identification of regulators for 40 selected yeast cell cycle regulated genes

Id	Target		best	$E \leqslant E_1$	$E \leqslant 1.1 * E_1$	$E \leqslant 1.2 * E_1$	Min(m)	Min(m) lin	E	
			m	m	m	m			Nonlin	Lin
1	YER150W	SPI1	3	4	5	8	4	2	0.0253	0.8339
2	YOR323C	PRO2	1	8	75	182	7	35	0.0010	0.0236
3	YKL177W	NA	0	0	0	5	7	3	0.0006	0.0277
4	YMR288W	HSH155	2	10	11	26	10	12	0.0019	0.0588
5	YMR316W	DIA1	4	15	29	40	21	1	0.0052	1.0992
6	YPL223C	GRE1	0	0	0	1	5	6	0.0017	0.0373
7	YPR035W	GLN1	2	2	2	10	2	6	0.0021	0.2907
8	YER003C	PMI40	1	2	3	4	1	11	0.0017	0.2779
9	YJL155C	FBP26	2	16	157	180	10	4	0.0003	0.0892
10	YMR145C	NDE1	0	0	3	10	4	16	0.0010	0.1342
11	YBR089W	NA	2	4	4	5	4	13	0.0577	1.4703
12	YDR285W	ZIP1	2	6	45	76	4	1	0.0274	1.8964
13	YFR057W	NA	0	0	13	46	8	4	0.0039	0.1206
14	YAL018C	NA	5	18	68	148	5	22	0.0003	0.1219
15	YOR383C	FIT3	2	2	2	6	2	14	0.0219	1.4964
16	YOR319W	HSH49	12	18	31	44	12	32	0.0801	4.7275
17	YOR264W	DSE3	7	7	16	20	7	7	0.0097	1.1955
18	YOL116W	MSN1	4	6	32	84	4	4	0.0045	0.1843
19	YGR269W	NA	0	0	1	5	2	1	0.0108	0.0778
20	YKL001C	MET14	4	13	23	27	3	1	0.0019	0.1988

As Criterion Decreases, Rate of False Positives Increases



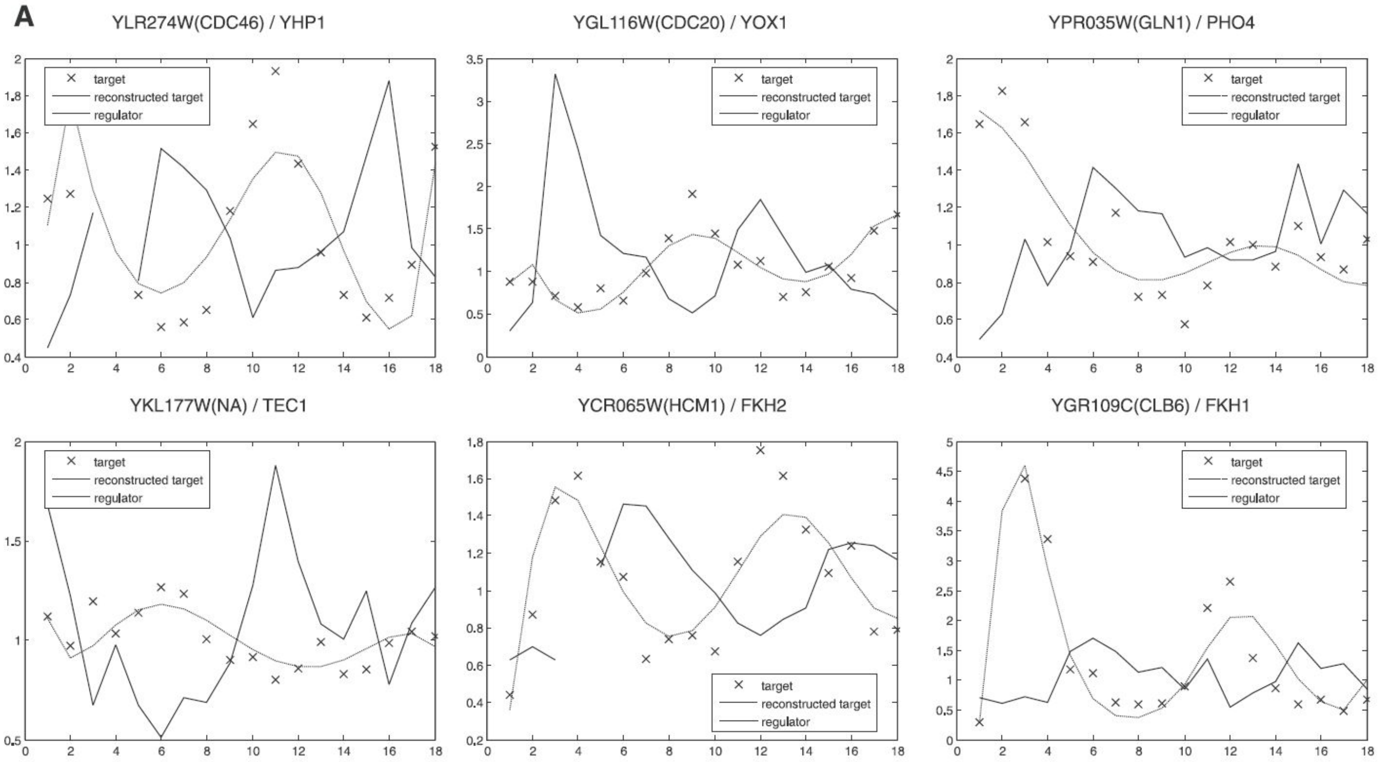
Id	Target		best m	$E \leqslant E_1$ m	$E \leqslant 1.1 * E_1$ m	$E \leqslant 1.2 * E_1$ m	Min(m)	Min(m) lin	E Nonlin	Lin
21	YDR146C	SWI5	0	0	0	1	4	12	0.0096	0.5309
22	YPL256C	CLN2	1	6	12	18	1	5	0.0253	1.2436
23	YJL187C	SWE1	1	2	3	6	1	4	0.0072	0.2139
24	YOR372C	NDD1	1	2	3	4	8	17	0.0062	0.1479
25	YLR274W	CDC46	2	7	5	6	7	7	0.0303	0.6388
26	YHR152W	SPO12	2	3	5	7	3	12	0.0012	0.3448
27	YCR065W	HCM1	2	6	6	8	6	16	0.0037	0.7056
28	YAL040C	CLN3	2	4	15	19	21	14	0.0105	0.7826
29	YDR224C	HTB1	1	3	3	3	3	2	0.0218	0.7135
30	YGL116W	CDC20	2	10	10	11	10	17	0.0050	0.5054
31	YPR119W	CLB2	4	7	9	13	8	21	0.0173	3.5841
32	YPL163C	SVS1	4	6	8	9	6	22	0.0360	7.7809
33	YLR210W	CLB4	0	0	0	0	15	3	0.0070	0.0858
34	YGR109C	CLB6	4	4	7	8	10	10	0.0922	5.9788
35	YBR010W	HHT1	0	0	1	1	7	5	0.0504	1.4994
36	YER111C	SWI4	2	21	24	27	1	1	0.0023	0.0000
37	YLR079W	SIC1	3	5	7	11	5	4	0.0384	0.5123
38	YER001W	MNN1	1	2	6	9	1	11	0.0193	3.5400
39	YDR225W	HTA1	1	4	4	3	4	9	0.0429	6.9192
40	YKL185W	ASH1	8	8	15	28	6	1	0.0173	0.0000
	% found		35	37.5	60	75	100	—	—	—

Vu, T. T., & Vohradsky, J. (2007)

Outline

- Goal: Identification of transcriptional regulators that control cell cycle regulated genes of Saccharomyces cerevisiae using non linear model
- Dynamic transcriptional control and optimization using least squares minimization
- Algorithm utilizes dynamic model of time continuous gene expression
- Method for determining correct regulators
- Trends in expression profiles for selected genes and their regulators
- Compared to linear model, nonlinear model gives better results in terms of correct identification of the regulator and better fit to gene expression profile of target.

Repressor and Target Gene Display Opposite Trends



Vu, T. T., & Vohradsky, J. (2007)

Activator and Target Gene Display Similar Trends

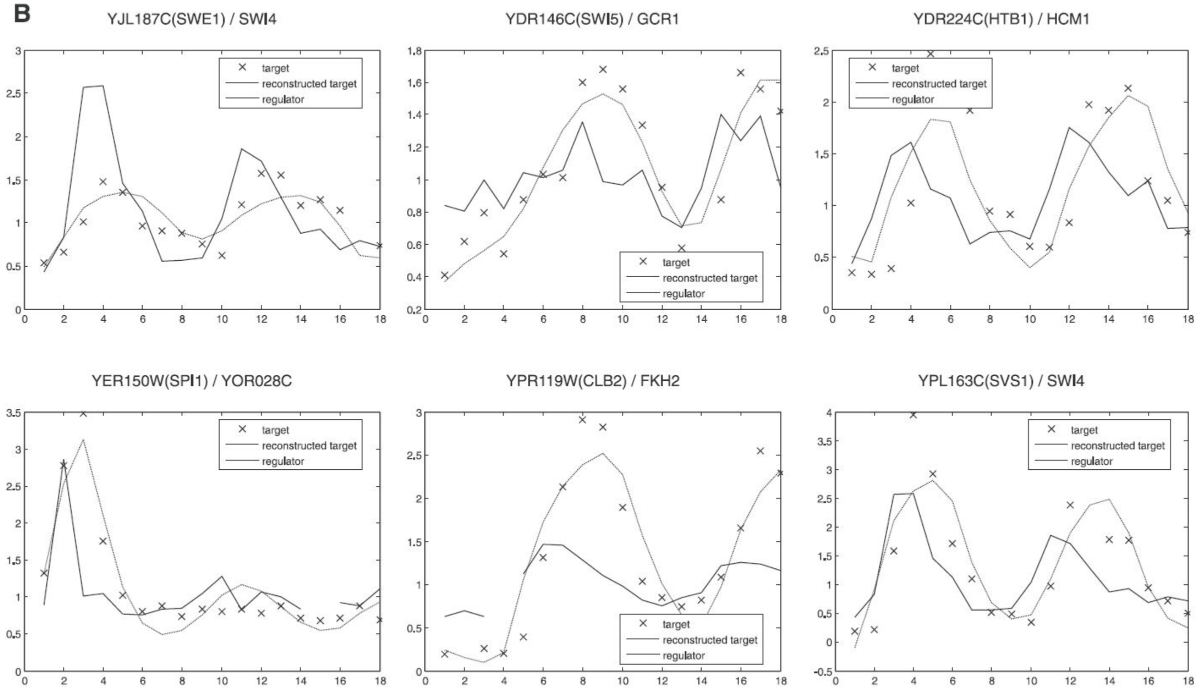


Figure 1. Expression profiles of 12 cell cycle regulated genes and their predicted regulators. (**A**) repressors, (**B**) activators. Horizontal axis—time points, vertical axis—expression relative to time point zero. Gene names in captions are arranged as target/regulator, symbols—target gene profile, dotted line—target gene profile fitted using the model, solid line—profile of the best fitting regulator (the lines are interrupted at the positions where the original data points were missing).

Vu, T. T., & Vohradsky, J. (2007)

Non Linear Model Give Better Results Compared to Linear Model

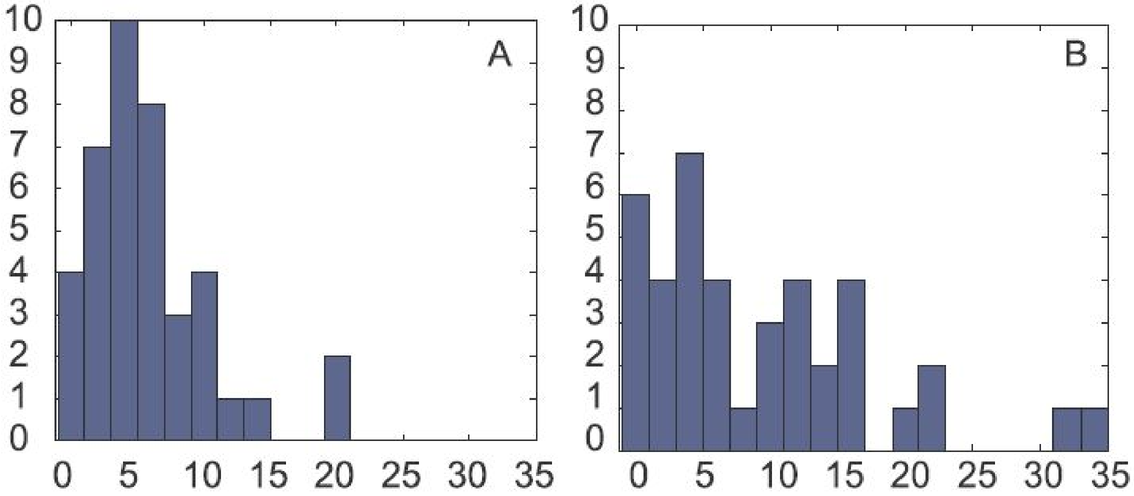


Figure 2. Histogram of distribution of the order of correctly identified regulators in the sorted list of potential regulators [columns Min(m) and Min(m) lin in Table 1], horizontal axis—the order in the sorted list. Regulators were sorted according to the error of approximation of the target gene expression profile (Equation 6). (**A**) Nonlinear model Equation 4, (**B**) linear model Equation 7.

Outline

- **Goal: Identification of transcriptional regulators that control cell cycle regulated genes of Saccharomyces cerevisiae using non linear model**
- **Dynamic transcriptional control and optimization using least squares minimization**
- **Algorithm utilizes dynamic model of time continuous gene expression**
- **Method for determining correct regulators**
- **Trends in expression profiles for selected genes and their regulators**
- **Compared to linear model, nonlinear model gives better results in terms of correct identification of the regulator and better fit to gene expression profile of target.**

Summary

- **Found that non linear model correctly identifies the regulators of target genes associated with the cell cycle in yeast and correctly determines their function (activator or repressor)**
- **Linear model gave lowest fit and lowest prediction ability when compared to both non linear model and model presented in Chen et. al (generalized linear model)**
- **When 3 models were compared (non linear, linear, Chen et. al) found that all 3 gave different results for the sets of genes**
- **The non linear model overall showed good accuracy and reasonable fit**
- **Since the model captures the behavior of transcriptional regulation/ provides info on influence of possible regulators and correctly predicts regulators, it can act as a useful tool in interpreting gene expression time series**
- **However, large scale network may require a large number of computations that may be unrealistic**