Validation of Ordinary Differential Equation Models for Parotid De-differentiation

JAEJIK KIM AND GRZEGORZ REMPALA
Department of Biostatistics, Medical College of Georgia, Augusta GA 30912, USA

ABSTRACT

Salivary glands are often damaged or destroyed by radiation therapy or surgery for head and neck cancers, or by advanced Sjogrenâs syndrome. In order to engineer or replace salivary glands, it is important to identify gene networks of the parotid acinar cells. Under this situation, Li(2010) proposed ordinary differential equation (ODE) models for three genes which have important roles in recovering salivary glands. However, since functional relationships between these three genes are unknown, we need to investigate the validation of these models and this is done by providing credibility bounds obtained from the Bayesian approach.

1 Introduction

Salivary glands are important for producing salivary proteins which contribute to host defense, lubrication, and digestion. However, salivary glands are often damaged or destroyed by radiation therapy or surgery for head and neck cancers, or by advanced Sjogrenâs syndrome. The serous acinar cell is the salivary cell type that is mos0t sensitive to radiation. Efforts to engineer or replace salivary glands are not likely to be fully successful due to our near-total lack of knowledge about the signaling and transcription factor pathways that direct terminal differentiation of acinar cells. Thus, it is important to define the major intracellular pathways of the nuclear program that causes terminal differentiation of the parotid acinar cell. Gene network discovery is a critical part to do this. Under this situation, Li(2010) proposed ordinary differential equation (ODE) models for three genes Mist1, PSP, and Amylase which are important roles in recovering salivary glands. However, in practice, functional relationships between these three genes are unknown. Thus, in this study, we investigate whether these models have the validity for given mRNA expression time series data using the Bayesian approach.

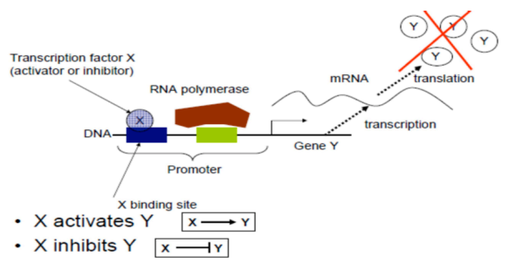
2 Methods

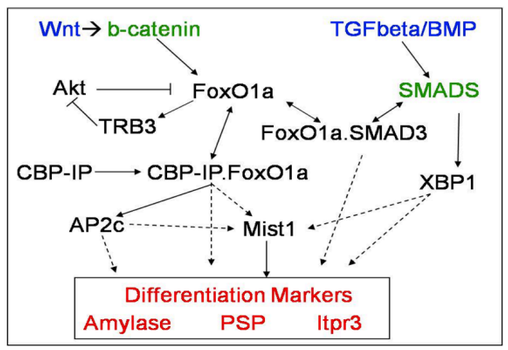
2.1 ODE models

The mechanism of the gene regulations can be described by

Rate of Change = Activation - Inhibition - Degradation.

Reciprocal regulation, feedback loop, and feed-forward loop involving transcription factors(TF) play critical roles.





Gene network of salivary gland (by D. Darling)

[2] hypothesized that the transcription is carried out by dimerization of Mist1 molecules and saturations of PSP and Amylase are attained when Mist1 is abundant. Under these hypotheses, the reactions and the ODE models proposed in [2] are as follows:

$$Mist1\,(M) \;\rightarrow\; PSP\,(P),\; Mist1\,(M) \;\rightarrow\; Amylase\,(A).$$

$$\frac{dA}{dt} = \frac{v_1 M^2}{v_2^2 + M^2} - v_3 A, \tag{1}$$

$$\frac{dM}{dt} = -v_4 M, \tag{2}$$

$$\frac{dP}{dt} = \frac{v_5 M^2}{v_6^2 + M^2} - v_7 P, \tag{3}$$

where $A(0) = 1$, $M(0) = 1$, and $P(0) = 1$, and where v_3, v_4, and v_7 are the degradation rates; v_1 and v_5 are the maximum reaction rate; and v_2 and v_6 are the half saturation points.

2.2 Experiments

Parotid glands from several rats are collected and pooled. The acinar cells are isolated from the fibroblasts. This single pool of acinar cells was then split into maximum 10 wells of 6-well plates, and cultured for 0, 1, 6, and 24 hours. At each time point, the cells in two different wells were are separately to make RNA samples. Thus, we have duplicated RNA samples at each time. These duplicates are cultured separately but originally come from the same pool of cells. In the next step, the expression of specific genes in the RNA samples are measured from RT-PCR assays. The result for each gene for each RNA samples are normalized to Gapdh gene which is an internal control that does not change. Then, for each gene and time, we have values for two biological replicates expressed relative to the initial value at 0 hour. These two biological replicates are averaged for each gene and time.
In the experiments, we assume that mRNA expressions cultured from different wells are independent. Thus, these average values of two biological replicates are independent for each gene and time. For the three genes of our interest, Amylase has three independent experiments, Mist1 is performed for seven independent experiments, and PSP has four independent experiments.

2.3 Validation

- **Time-course data**:
  - $y_{ij}(\mathbf{t})$, $i = A, M, P$, $j = 1, \ldots, r_i$: $k \times 1$ vectors for Amylase, Mist1, and PSP (j-th experiment for gene i).
  - Vector of time points: $\mathbf{t} = (t_1, \ldots, t_K)'$, where $K = 3$ and $t_1 = 1$, $t_2 = 6$, and $t_3 = 24$.
  - The number of experiments: $r_A = 3$, $r_M = 7$, and $r_P = 4$.
- **Model**:

$$y_{ij}^L(\mathbf{t}) = y_{ij}^M(\mathbf{t}|\mathbf{v}_i) + b_i(\mathbf{t}) + \epsilon_{ij}(\mathbf{t}),\; i = A, M, P,\; j = 1, \ldots, r_i, \tag{4}$$

  - $y_{ij}^L(\mathbf{t})$: the vector of the actual measurements.
  - $y_{ij}^M(\mathbf{t}|\mathbf{v}_i)$: outputs from the ODE models in (1), (2), and (3).
  - $\mathbf{v}_i$, $i = A, M, P$: the vector of parameters of the ODE models for each gene (i.e., $\mathbf{v}_A = (v_1, v_2, v_3)'$, $\mathbf{v}_M = v_4$, and $\mathbf{v}_P = (v_5, v_6, v_7)'$).
  - $b_i(\mathbf{t})$ ($\equiv \mathbf{b}_i$): the model bias.
  - $\epsilon_{ij}(\mathbf{t})$: the vector of measurement error.
- **Likelihood function**:

$$L(y_i^L(\mathbf{t})|\mathbf{v}_i, \mathbf{b}_i) \sim MVN(\mu_i,\; \delta_i \mathbf{I}), \tag{5}$$

  - $\mu_i = (f_i(t_1, \mathbf{v}_i) + b_i(t_1), \ldots, f_i(t_k, \mathbf{v}_i) + b_i(t_K))'$, where $f_i(\cdot)$ is the solution of the ODE models defined in (1), (2), and (3).

- **Prior distribution**:
  - v_i's have independent gamma distributions.
  - $\mathbf{b}_i \sim MVN(\mathbf{0}, \mathbf{C})$, where $\mathbf{C} = [\exp\{-(t_i - t_j)^2\}]_{mn}$.
- **Posterior distribution**:

$$p(\mathbf{v}_A, \mathbf{v}_M, \mathbf{v}_P, \mathbf{b}_A, \mathbf{b}_M, \mathbf{b}_p | y_A^L(\mathbf{t}), y_M^L(\mathbf{t}), y_P^L(\mathbf{t})) \propto \prod_{i=\{A,M,P\}} L(y_i^L(\mathbf{t})|\mathbf{v}_i, \mathbf{b}_i) p(\mathbf{v}_i, \mathbf{b}_i).$$

- **MCMC analysis**:
  - $\mathbf{v}_{il}$ and $\mathbf{b}_{il}$, $i = A, M, P$, $l = 1, \ldots, n$: n posterior samples.
  - Posterior mean for $\mathbf{v}_i$:

$$\hat{\mathbf{v}}_i = \frac{1}{n}\sum_{l=1}^{n} \mathbf{v}_{il},\; i = A, M, P. \tag{6}$$

  - Posterior mean for $\mathbf{b}_i$:

$$\hat{\mathbf{b}_i} = \frac{1}{n}\sum_{l=1}^{n} \mathbf{b}_{il},\; i = A, M, P. \tag{7}$$

  - Bias-corrected prediction:

$$\hat{y}_i^L(\mathbf{t}) = \frac{1}{n}\sum_{l=1}^{n} \left[y_i^M(\mathbf{t}, \mathbf{v}_{il}) + \mathbf{b}_{il}\right],\; i = A, M, P. \tag{8}$$

  - Pure model prediction:

$$\hat{\mathbf{b}}_i^{\hat{\mathbf{v}}_i} = \hat{y}_i^L(\mathbf{t}) - y_i^M(\mathbf{t}, \hat{\mathbf{v}}_i),\; i = A, M, P. \tag{9}$$

  - Variance of the bias-corrected predictor (8):

$$V\{\hat{y}_i^L(\mathbf{t})\} = \frac{1}{n}\sum_{l=1}^{n} \left[y_i^M(\mathbf{t}, \mathbf{v}_{il}) + \mathbf{b}_{il} - \hat{y}_i^L(\mathbf{t})\right]^2,\; i = A, M, P. \tag{10}$$

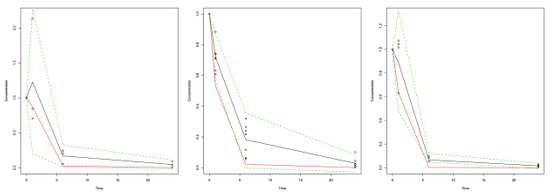
  - $(1-\alpha)\%$ credibility bound: $\alpha/2$ and $(1-\alpha/2)\%$ quantiles for $(y_i^M(\mathbf{t}, \mathbf{v}_{il} + \mathbf{b}_{il})$.
- **Sensitivity analysis**:

$$sens_{v_m}(t, \hat{\mathbf{v}}_i) = \left.\frac{f_i^{v_m}(t, \mathbf{v}_i)}{f_i(t, \mathbf{v}_i)}\right|_{\mathbf{v}_i = \hat{\mathbf{v}}_i},\; m = 1, \ldots, 7,\; i = A, M, P. \tag{11}$$

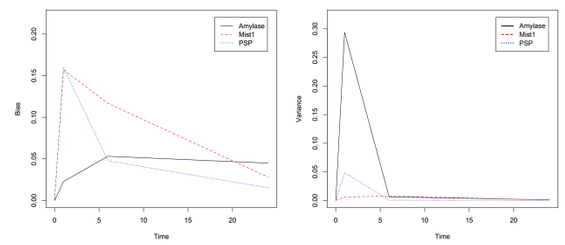
  - $f_i^{v_m}(t, \mathbf{v}_i) = \partial f_i(t, \mathbf{v}_i)/\partial v_m$: gradient of $f_i(t, \mathbf{v}_i)$ in the direction of v_m.

3 Results

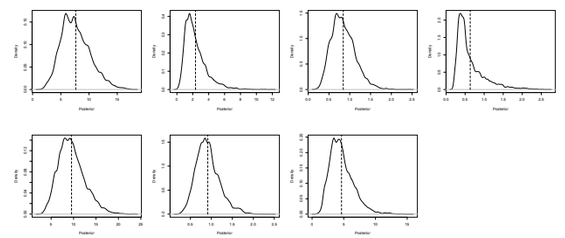
It is not easy to obtain the convergence of posterior distributions with high-dimension in the MCMC analysis. In our case, we deal with 16 parameters (16 dimensions). For the given data, we could not obtain the convergence of all 16 parameters simultaneously in the MCMC procedure. Thus, we fixed bias terms of PSP at 1 and 6 hours (i.e., $b_P(t_1)$ and $b_P(t_2)$) by the posterior means obtained when the other parameters are fixed. For fixed $b_P(t_1)$ and $b_P(t_2)$, we attained the convergence of the posterior distributions for the other 14 parameters.



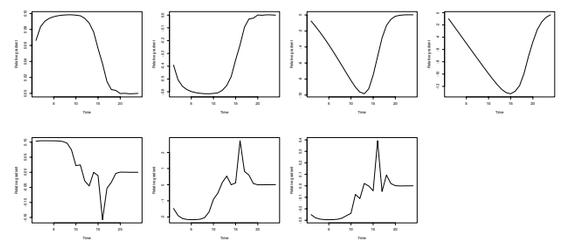
Actual data (circle) vs pure model prediction $y_i^M(\mathbf{t}, \mathbf{v}_i)$ (red line) and bias corrected model prediction $\hat{y}_i^L(\mathbf{t})$ (black line) along with 95% credibility bounds (green dotted line); (**Left panel**) Amylase, (**Center panel**) Mist1, (**Right panel**) PSP.



Bias (**Left panel**) and variance (**Right panel**) of the model prediction for data via (9) and (10).



The posterior distributions of the parameters $v_1, \ldots, v_7$ of the ODE models.



The relative gradient of the ODE models, that is, $sens_{v_m}(t, \hat{\mathbf{v}}_i)$, $m = 1, \ldots, 7$ given by (11).

4 Conclusion

The ODE models proposed by Li(2010) hypothesized that the transcription is carried out by dimerization of Mist1 molecules and saturations of PSP and Amylase are attained when Mist1 is abundant. In this study, for the observed data, we test these hypotheses by using the Bayesian approach, and the results from the MCMC analysis support the hypotheses of these ODE models.

References

[1] M.J. Bayarri, J.O. Berger, R. Paulo, and J. Sacks, A framework for validation of computer models, Technometrics 49 (2007) 138–154.

[2] J. Li, ODE models of parotid de-differentiation, Presentation at the University of Louisville (2010).

[3] G.A. Rempala, K.S. Ramos, T. Kalbfleisch, I. Teneng, Validation of a mathematical model of gene transcription in aggregated cellular systems: application to L1 retrotransposition, J. Comp. Biol. 47 (2007) 339–349.