



INFERENCIA DE PARÁMETROS EN UNA RED DE REGULACIÓN GENÉTICA

Brenda Aide Peña Cantu
J. Arturo Berrones Santos
Edgar Jiménez Peña
Francisco Javier Almaguer Martínez

Resumen

En el presente trabajo se abordan diferentes tipos de metodologías para la inferencia de parámetros en redes de interacción genética. Se presentan resultados en base al estudio realizado en una red artificial específica conocida como el *Represilador*. Primeramente de métodos de muestreo de datos empleados. Posteriormente se presentan diversos tipos de suavizado y su desempeño en base al funcionamiento de la red estudiada. El problema es motivado por un caso de estudio real en líneas celulares de cáncer al cuál se pretende extender el presente análisis.

Palabras clave: biología de sistemas, red de regulación genética, *Represilador*, simulación, problema inverso.

Introducción

En la actualidad los problemas relacionados con la salud son de gran interés de estudio. La biología de sistemas es una rama joven de la ciencia que integra múltiples disciplinas como biología, química, física, medicina, ingeniería, matemáticas, informática, etc. Esto con el fin de obtener un mayor entendimiento de los sistemas vivos y de sus procesos biológicos (López, Ruiz-Romero, y Vega, 2007).

Los procesos de regulación genética se encuentran inmersos entre los muchos mecanismos estudiados por la biología de sistemas. El estudio de estos procesos es de inmensa importancia ya que suceden dentro de cada organismo vivo y facilita una gran cantidad de información acerca de las interacciones entre genes y proteínas para realizar posibles formulaciones de hipótesis biológicas (Xu, Venayagamoorthy y Wunsch II, 2007).

La inferencia de procesos de regulación genética es estudiada mediante redes complejas con el fin de entender diversas tareas bioquímicas como procesos celulares, funciones de genes y además comprender las relaciones e interacciones entre genes (De-Jong, 2002).

Una red de regulación genética consiste en un conjunto de DNA, RNA, proteínas y otras moléculas y su descripción de mecanismos de regulación mediante estos componentes (Xu, Venayagamoorthy y Wunsch II, 2007).

El problema presentado en este trabajo es inspirado en una red de regulación genética artificial conocida en la literatura como *Represilador* (Elowitz y Leibler, 2000), (Buse, Pérez y Kuznetsov, 2010). En esta red se desea conocer los parámetros involucrados en la interacción de los genes con el objetivo de describir su comportamiento. En la Figura 1 podemos observar las ecuaciones diferenciales asociadas a la red descrita. Donde u , v y w representan la concentración de cada uno de los genes descritos. Las m 's corresponden a la concentración de RNA mensajero relacionadas a los genes. Además de los parámetros α , α_0 y n los cuales representan el número de copias por célula de proteína producida, la relación entre velocidad de desintegración de proteínas y el coeficiente de Hill respectivamente.

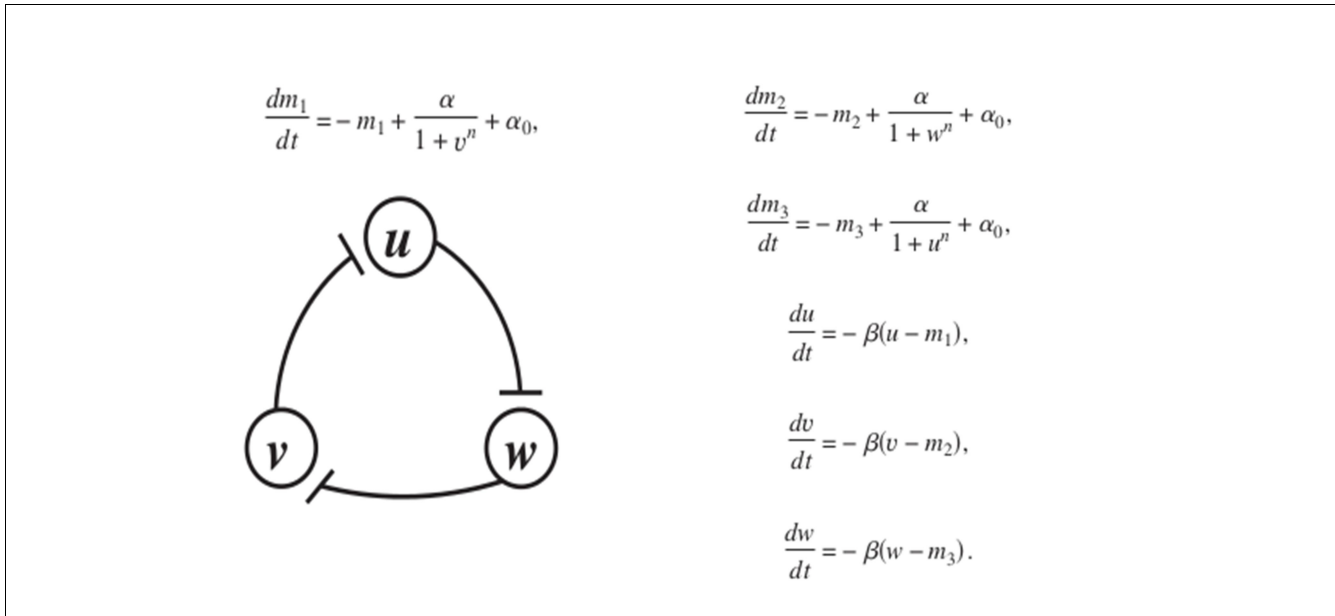


Figura 1. Conjunto de Ecuaciones Diferenciales del *Represilador*.

Para esto se emplearon técnicas de simulación con el fin de reproducir el comportamiento de la red antes descrita.

Justificación

Realizar este tipo de estudio es de un impacto significativo, ya que tiene múltiples aplicaciones en áreas como la medicina y la biología, por lo que se considera un desarrollo científico de gran importancia.

Metodología

Logrando la simulación de la red se procedió a la etapa de inferir los parámetros de ésta mediante el uso del problema inverso. El problema inverso es aquel en donde los valores de algunos parámetros del modelo deben ser obtenidos de los datos observados (Engl, et. al., 2009).

El estudio realizado en este trabajo consistió en utilizar diversos tipos de muestreos para inferir valores en la red de regulación genética artificial. Posteriormente se realizaron diferentes tipos de suavización a los datos obtenidos mediante los muestreos con el fin de tener un mejor ajuste en los valores de los parámetros.

Es necesario tomar en cuenta el número de datos que se desea muestrear, además del tipo de muestreo deseado. En este estudio se tienen dos, aleatorio y sistemático. Además se considera un error de medición, que en un caso real se le relaciona con fallas humanas y perturbaciones en el entorno del sistema. Las suavizaciones se aplicaron a los datos muestreados de forma sistemática. Se utilizaron tres tipos de procedimientos de suavización como aproximaciones por regresiones de polinomios locales, estimaciones de suavizaciones cúbicas y funciones de estimación mediante polinomios locales. Integrados en el software libre R (R Core Team, 2012).

Resultados

En esta sección del trabajo presentamos algunos de los resultados obtenidos en el estudio de la red de regulación genética *Represilador*.

En la Figura 2 se muestran las simulaciones de las series de interacción entre los genes. Donde en el lado izquierdo se observan las 3 concentraciones de proteínas por células correspondientes a cada uno de los genes involucrados. A su mismo en el lado derecho se muestra la concentración del RNA mensajero.

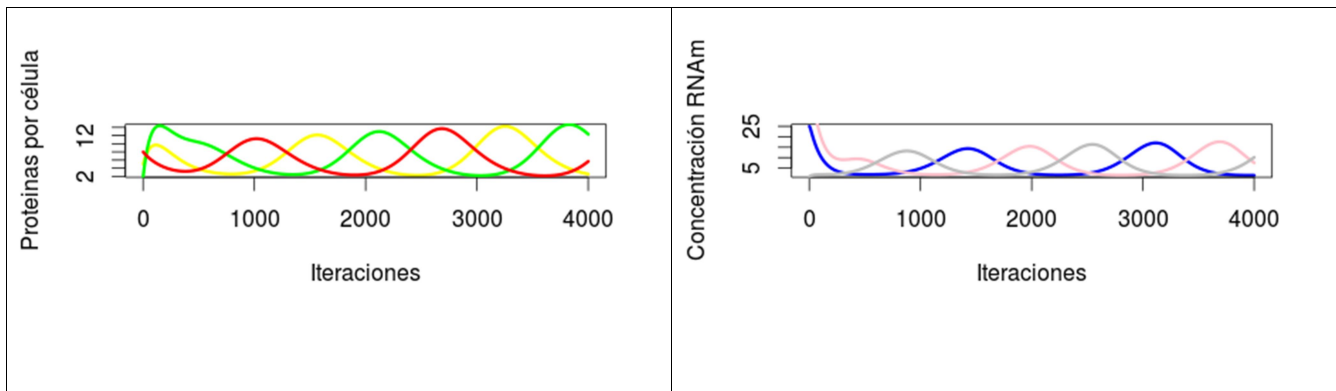


Figura 2. Simulación del *Represilador*.

Se estudiaron 3 tamaños de muestra, 200, 50 y 10. Con un error (ruido de medición) del 1 % para cada caso. En la Figura 3 se presentan los datos obtenidos usando un tamaño de muestra de 200 datos y un error de medición del 1%. En el lado izquierdo se observa las series obtenidas mediante un muestreo aleatorio y el lado derecho muestra la serie de datos con muestreo sistemático.

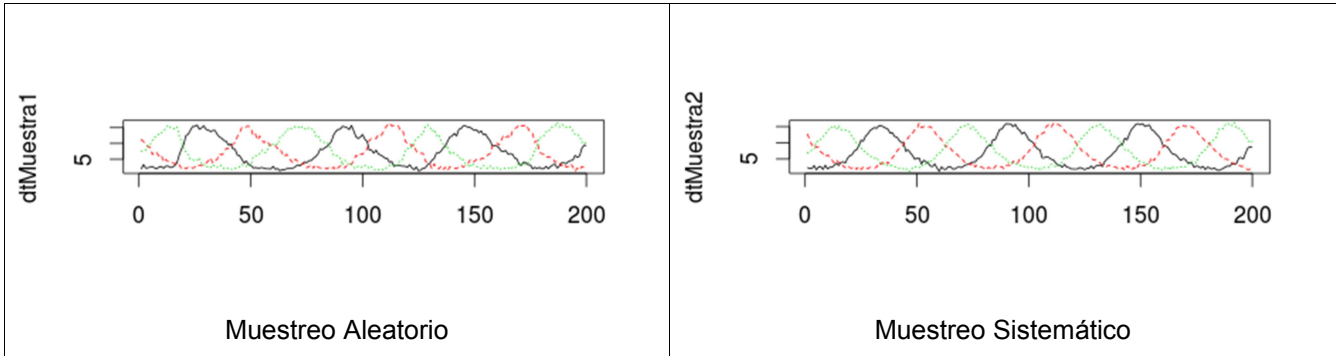


Figura 3. Muestreos con tamaño 200 y 1% de error.

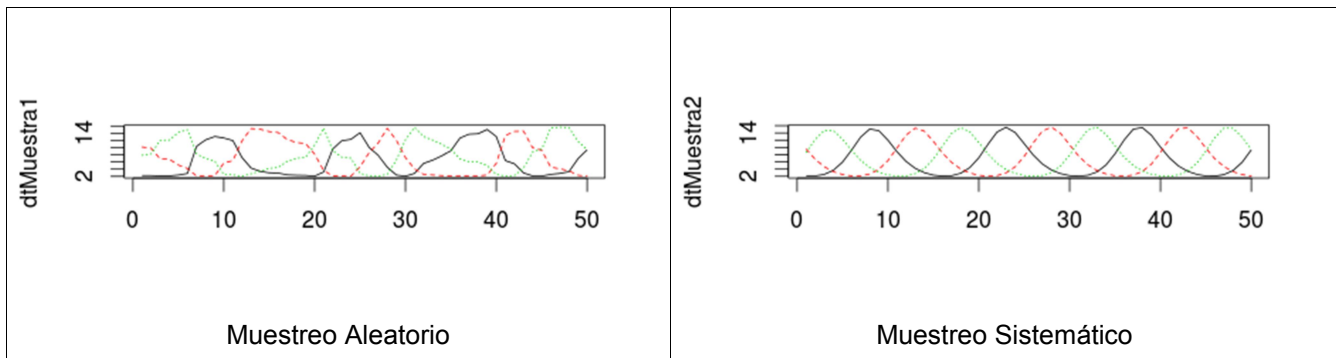


Figura 4. Muestreos con tamaño 50 y 1% de error.

En la Figura 4 se observan los datos obtenidos para el muestreo con 50 datos y 1% de error. Donde los datos están relacionados de la misma manera que la Figura 3.

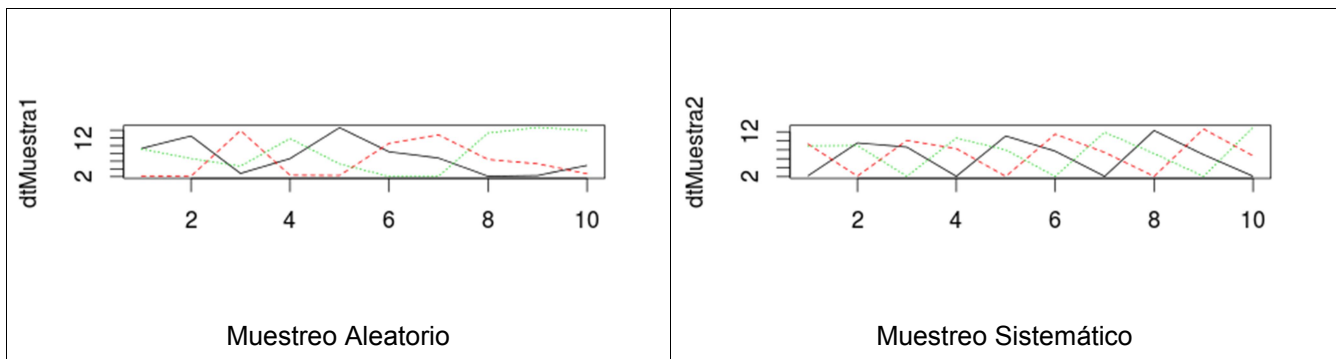


Figura 5. Muestreos con tamaño 10 y 1% de error.

Por último en la Figura 5 observamos los datos muestreados correspondientes a tamaño 10 y 1% de error. Donde observamos de igual manera los dos tipos de muestreos empleados.

Aplicando los tipos de suavización a los datos muestreados de forma sistemática se pudo construir la Figura 6 donde se resume el desempeño de las 3. En el eje y de la gráfica se encuentra el logaritmo del error entre la muestra y la serie original. El eje x corresponde a la cantidad de datos. Por último aparecen los 3 tipos de suavización; loess en color rojo, spline en verde y kernel en azul.

Lo que se observa claramente en la Figura 6 es que el desempeño de los tipos de suavización mejora paulatinamente conforme el número de datos aumenta. Y además se muestran las líneas trazadas de cada suavización empleada.

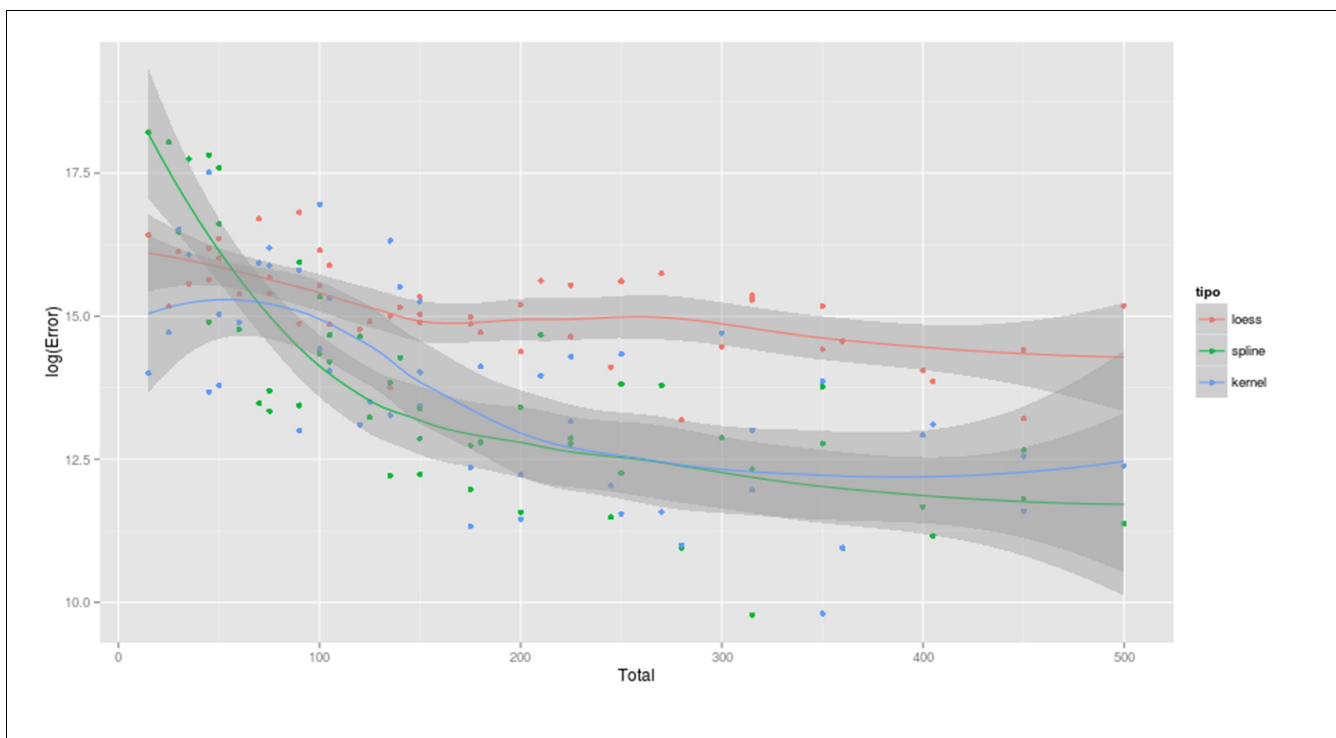


Figura 6. Resumen de desempeño de suavizaciones.

Conclusiones

Como se ha mostrado en los resultados anteriores los métodos de muestreo logran recuperar la estructura oscilatoria de la red de regulación genética estudiada en el presente trabajo. Al mismo tiempo el tamaño de los datos juega un papel importante, ya que la escasez de datos es muy común en casos reales. Por lo que se concluye que se debe mejorar ambos métodos, tanto los muestreos como las suavizaciones. Además de incluir metodologías estadísticas con el fin de que se alcance una robustez en la inferencia de los datos. De tal modo que sea posible extenderlo a redes de interacción genética reales y concretas.

Bibliografía

Buse, O., Pérez, R. y Kuznetsov, A. Dynamical Properties of the Repressilator Model. *Physical Review*. 81(7):1-7, 2010.

De-Jong, H. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. *Journal of Computational Biology*. 9(1):67-103,2002.

Elowitz, M., B. y Leibler, S. A Synthetic Oscillatory Network of Transcriptional Regulators. *Nature*. 403(2000):335-338, 2000.

Engl, H., W., Flamm, C., Kugler, P., Lu, J., Muller, S. y Schuster, P. Inverse Problems in Systems Biology. *Top Science*. 25(2009):1-51, 2009.

López, M., Ruiz-Romero, G., y Vega, M. (2007). *Biología de Sistemas*. FUAM,(GEN-ES07003). Madrid:Genoma España.

R Core Team (2012). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.

Xu, R., Venayagamoorthy, G., K. y Wunsch II, D., C. Modeling of Gene Regulatory Networks with Hybrid Differential Evolution and Particle Swarm Optimization. *Science Direct*. 20(2007):917-927, 2007.