

TITRE (Français)	Identification de réseaux de régulation génique à partir de données d'expression : une approche basée sur les modèles affines par morceaux.
TITRE (English)	Identification of genetic regulatory networks based on expression data: an approach based on piecewise affine models.
AUTEUR	Samuel Drulhe
UNIVERSITE	Université Joseph-Fourier - Grenoble I
DATE	12/09/08
LABORATOIRE	INRIA
DIRECTION DE THESE	Hidde de Jong (INRIA), Giancarlo Ferrari-Trecate (Uni. Pavia)
PARRAINAGE	

Résumé en Français :

Les progrès récents des techniques expérimentales biologiques ont conduit à la production d'une énorme quantité de données sur le comportement dynamique des réseaux de régulation génique (RRG). Nous présentons une approche pour l'identification des modèles affines-par-morceaux (APM) de RRGs à partir de données expérimentales. Ces modèles reposent sur l'hypothèse que la régulation survient au niveau de la synthèse et de la dégradation des produits de l'expression des gènes : les paramètres cinétiques sont supposés être constants jusqu'à ce que la concentration d'une protéine régulatrice franchisse un seuil de transition.

La méthode que nous présentons se concentre sur le problème de la détection des transitions entre les différents modes dynamiques à partir des données d'expression génique et sur la reconstruction des seuils de transition associés avec les interactions régulatrices. En particulier, notre méthode prend en considération les contraintes géométriques spécifiques aux modèles APM de RRGs. Une telle méthode d'identification est conçue pour des systèmes à erreur sur la sortie où les observations sont des séries temporelles de mesures bruitées de niveaux de concentration à l'intérieur d'une cellule.

Les données sont d'abord classées en modes dans lesquels le comportement dynamique est considéré comme étant complètement décrit par une équation différentielle linéaire. À partir de la classification résultante, une technique de reconnaissance de forme est utilisée pour reconstruire toutes les combinaisons de seuils de transition qui sont cohérentes avec les données mesurées. Pour chaque combinaison de seuils, il est alors possible de fournir un réseau de régulation et les paramètres dynamiques de chaque mode.

Les performances de notre approche ont été analysées en utilisant des données artificielles simulées pour un modèle simplifié de la réponse à un manque de carbone pour la bactérie *Escherichia coli*. En particulier, nous avons évalué l'influence du niveau du bruit et du pas d'échantillonnage sur les systèmes identifiés. Nos résultats montrent que la méthode, en association avec des séries temporelles de mesures suffisamment précises, lesquelles peuvent être obtenues avec des systèmes à gène rapporteur, permettent une identification quantitative de modèles APM de RRGs.

Résumé en anglais :

Recent advances of experimental techniques in biology have led to the production of enormous amounts of data on the dynamics of genetic regulatory networks (GRN). We present an approach for the identification of piecewise affine (PWA) models of GRNs from experimental data. These models rely on the assumption that regulation happens at the level of synthesis and degradation of gene expression products: kinetic parameters are considered as being constant until a regulating protein concentration crosses a switching threshold.

The method we present focuses on the problem of detecting switches among different modes of operation in gene expression data and on the reconstruction of switching thresholds associated with regulatory interactions. In particular, our method takes into account geometric constraints specific to PWA models of GRNs. Such an identification method is designed for output-error systems where the observations are noisy time-series measurements of concentration levels inside a cell.

Data points are first classified into modes on which the dynamical behavior is assumed to be fully described by a linear differential equation. From the resulting classification, a pattern recognition technique is used to reconstruct all combinations of switching thresholds that are consistent with measured data.

For each combination of thresholds, it is then possible to provide an identified regulatory network and the dynamical parameters of each mode of operation.

*The performance of our approach has been analyzed using synthetic data simulated from a simplified model of the carbon starvation response in the bacterium *Escherichia coli*.*

In particular, we evaluated the impact of the noise level and sampling time on the identified systems.

Our results show that the method, coupled with sufficiently precise time-series data, which can be obtained from gene reporter systems, enables a quantitative identification of piecewise affine models of genetic regulatory networks.

<http://tel.archives-ouvertes.fr/tel-00380505/fr/>

oai:tel.archives-ouvertes.fr:tel-00380505