



Review Article

A glance at the applications of Singular Spectrum Analysis in gene expression data

Hossein Hassani^{a,b,*}, Zara Ghodsi^b^a Institute for International Energy Studies (I.I.E.S), 1967743711 Tehran, Iran^b The Statistical Research Centre, Bournemouth University, UK

ARTICLE INFO

Article history:

Received 7 January 2015

Received in revised form 12 March 2015

Accepted 8 April 2015

Available online 29 May 2015

Keywords:

Singular Spectrum Analysis

Filtering

Signal extraction

ABSTRACT

In recent years Singular Spectrum Analysis (SSA) has been used to solve many biomedical issues and is currently accepted as a potential technique in quantitative genetics studies. Presented in this article is a review of recent published genetics studies which have taken advantage of SSA. Since Singular Value Decomposition (SVD) is an important stage of this technique which can also be used as an independent analytical method in gene expression data, we also briefly touch upon some areas of the application of SVD. The review finds that at present, the most prominent area of applying SSA in genetics is filtering and signal extraction, which proves that SSA can be considered as a valuable aid and promising method for genetics analysis.

© 2015 The Authors. Published by Elsevier GmbH. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).

Contents

1. Introduction.....	17
2. Signal extraction and filtering.....	18
2.1. SSA based on minimum variance.....	18
2.2. SSA combined with AR model.....	19
2.3. SVD.....	20
3. Application of two dimensional SSA.....	20
4. Conclusion.....	20
References.....	21

1. Introduction

Nowadays there exists a large amount of datasets in the field of genetics and expression measurement and there are many different methods and techniques for analysing these datasets [1–3]. However, it has been widely accepted that the major difficulties in working with such data no longer is the validity of expression measurements, but the reliability of inferences from the data as the achieved data is difficult to understand without the use of proper analytical tools [4] and if research result is obtained from an inappropriate model it can never be translated into a valid scientific context [5].

Historically, such data have been analysed using parametric methods [6,7]. However, constraining pre-assumptions needed

for parametric approaches led towards the growing popularity of nonparametric methods. Recently it has been concluded that nonparametric techniques can be used as an alternative approach for analysing genetics data because of their inherent nature [8], and accordingly the applications in biomedical and genetics fields have expanded.

Among many non-parametric methods, Singular Spectrum Analysis (SSA) is a relatively new approach which has proven to be very successful. SSA has already transformed itself into a standard tool in the analysis of biomedical, mathematical, geometrical and several other time series [9–11] and recently it has also been applied in genetics which has illustrated its strong potential for such studies [12,13].

The emergence of SSA is usually associated with the work by Broomhead in 1986 [14]. However, the ideas of SSA were independently developed in Russia and in several groups in the UK and USA. Since after, several papers on the methodology and applications of SSA have been published (see, for example, [15,16]). An introduction to this technique can be identified in the paper by Elsner and

* Corresponding author at: The Statistical Research Centre, Bournemouth University, UK. Tel.: +44 7703367456; fax: +44 7703367456.

E-mail address: hhassani@bournemouth.ac.uk (H. Hassani).

Tsonis [17] and a comprehensive description with several examples of theoretical and practical aspects of SSA can also be found [8,18].

The main advantages of the SSA technique in the field of genetics can be attributed to its signal extraction and filtering capabilities [19], batch processing of a set of similar series [20] and derivation of an analytical formula of the signal [21]. The application of SSA for noise reduction in microarrays, and signal extraction in gene expression data has received more interest. The reason underlying the significant interest in the SSA technique's filtering capabilities are due to the fact that genetics data is often characterised by the existence of considerable noise, filtering this noisy data is considered as one of the most arduous tasks when analysing genetics data [22,23].

For example, microarray is a very useful method for acquiring quantitative data in genetics and researchers today are conducting most of their studies using this method. The main advantage of microarray is the capability of studying thousands of genes simultaneously. However, microarray data usually contains a high level of noise, which can reduce the performance of the results [24].

This article categorises and summarises almost all recently published articles associated with the application of SSA in genetics.

Presented below, is a short description of SSA technique in doing so we mainly follow [9] where a more detailed description is made available. Moreover, the R package for this technique including decomposition, forecasting and gap filling for univariate and multivariate time series can be downloaded via [25]

Consider a set of genetics observations in a series of $Y_N = (y_1, \dots, y_N)$ with length of N . After choosing a window length L where $(2 \leq L \leq N-1)$, we construct the L -lagged vectors $X_j = (y_j, \dots, y_{L+j-1})^T$, $j=1, \dots, K$ where $K=N-L+1$. Define the matrix $\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = (X_1, \dots, X_K)$. Now \mathbf{X} is our multivariate data with L characteristics and K observations. The columns X_j of \mathbf{X} , are the vectors, positioned in an L -dimensional space \mathbb{R}^L . Define the matrix $\mathbf{X}\mathbf{X}^T$: SVD of $\mathbf{X}\mathbf{X}^T$ gives us the collections of L eigenvalues $(\lambda_1 \geq \dots \geq \lambda_L \geq 0)$ and the corresponding eigenvectors $U_1 \dots U_L$ where U_i is the normalised eigenvector corresponding to the eigenvalue $\lambda_i (i=1, \dots, L)$. A group of r (with $1 \leq r < L$) eigenvectors determines an r -dimensional hyperplane in the L -dimensional space \mathbb{R}^L of vectors X_j . By choosing the first r eigenvectors U_1, \dots, U_r , then the squared L_2 -distance between this projection and \mathbf{X} is equal to $\sum_{j=r+1}^L \lambda_j$. Based on the SSA process, the L -dimensional data are projected onto this r -dimensional subspace and the final diagonal averaging gives us an appropriate approximation of the first one dimensional series.

The remainder of this paper is organised as follows. In the following section we present a review of papers involving the application of SSA and SVD¹ on signal extraction and noise reduction. The SSA based on minimum variance and a hybrid modelling approach combining SSA and autoregressive (AR) model are also discussed in depth in that section. Section 3 provides theoretical developments leading to what is termed as "two-dimensional SSA" and the paper ends with some conclusions in Section 4.

2. Signal extraction and filtering

In this section we identify existing applications of SSA for signal extraction and noise filtering in genetics.

The first such application is reported in 2006 where SSA was used for signal extraction of *Drosophila melanogaster*'s gene

expression profile [19]. The idea of using SSA for signal extraction was then followed in an approximately similar study in [26] which led to an improved result. By 2008 a more technical study conducted on the methodology of signal extraction from the noisy *Bicoid* (*Bcd*) protein profile in *Drosophila melanogaster* was presented in [21]. The problem under investigation in that study was complicated by two facts: (i) the data contained outliers and (ii) that the data was exceedingly noisy and the noise consisted of an unknown structure. The author examined two approaches for reconstructing signal more precisely: the use of small window length and improvements to separability by adding a constant to the series.

In addition, the activation of the *hunchback* (*hb*) gene in response to different concentrations of *Bcd* gradient was studied in [27] and SSA was applied for filtering two kinds of noise; experimental noise and the noise caused by variability in nuclear order [27].

2.1. SSA based on minimum variance

Recently, a modified version of SSA was examined for filtering and extracting the *bcd* gene expression signal [28] and the results illustrated that SSA based on minimum variance can significantly outperform the previous methods used for filtering noisy *Bcd* [28].

SSA based on minimum variance mainly relies on the concept that by dividing the given noisy time series into the mutually orthogonal noise and signal+noise components, an enhanced estimation of the signal can be achieved. Thus, after performing SVD, by adapting the weights for different obtained singular components, an estimation of the Hankel matrix \mathbf{X} , will be achieved which in turn corresponds to a filtered series.

A short description of the SSA based on minimum variance is given below. For more details, see [28,29].

Let us begin with the Singular Value Decomposition (SVD) of the trajectory matrix \mathbf{X} :

$$\mathbf{X} = \mathbf{U}(\mathbf{W}\Sigma)\mathbf{V}^T, \quad (1)$$

where \mathbf{W} is the diagonal matrix of the weights to be determined. The SVD of the matrix \mathbf{X} can be written as:

$$\mathbf{X} = [\mathbf{U}_1 \quad \mathbf{U}_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix}, \quad (2)$$

where $\mathbf{U}_1 \in \mathbb{R}^{L \times r}$, $\Sigma_1 \in \mathbb{R}^{r \times r}$ and $\mathbf{V}_1 \in \mathbb{R}^{K \times r}$.

Now, the issue is in selecting the weight matrix \mathbf{W} . If we represent the SVD of the Hankel matrix related to the signal as \mathbf{S} , by considering different criteria in choosing this matrix, different estimation of \mathbf{S} can be achieved. The LS Estimate of \mathbf{S} is the current widely used approach in selecting the weight matrix \mathbf{W} . This approach is based on the idea of removing the noise subspace but keeping the noisy signal uncorrelated in the signal+noise subspace. However, the accuracy of this estimator is mainly dependent on the estimation of the signal rank r since selecting singular values in LS follows a binary approach. Although in using this estimator, considering any assumptions is not needed.

In MV Estimate of \mathbf{S} proposed by Hassani in [29], removing the noise components in the signal + noise subspace has been improved considerably. However, to obtain the MV estimate considering some assumptions is essential (for more details see [29]).

Let us now consider the weight matrix \mathbf{W} of the LS and MV estimates:

$$\begin{aligned} \hat{\mathbf{S}}_{LS} &= \mathbf{U}_1(\mathbf{W}_{LS}\Sigma_1)\mathbf{V}_1^T \\ \hat{\mathbf{S}}_{MV} &= \mathbf{U}_1(\mathbf{W}_{MV}\Sigma_1)\mathbf{V}_1^T, \end{aligned} \quad (3)$$

¹ The SVD algorithms used in this paper are either based on Hankel or Teoplitz matrix.

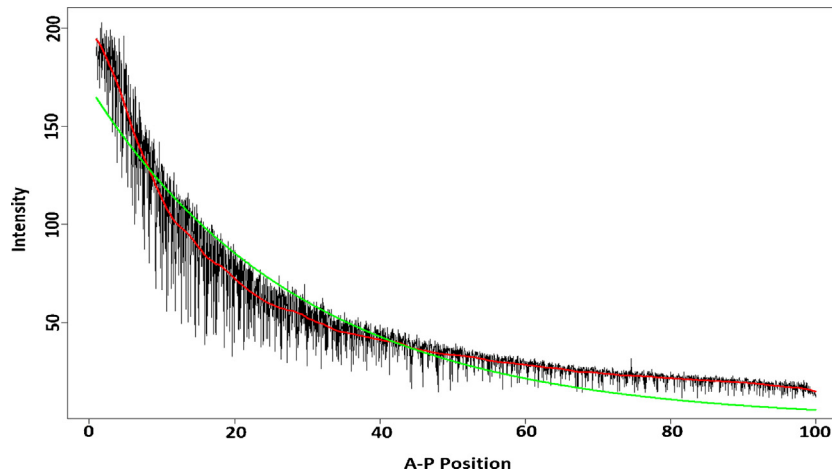


Fig. 1. Temporal dynamics of *Bcd* expression, cleavage cycle 14(6). Red and green lines show the trend extracted by SSA_{MV} and SDD [28]. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of the article.)

where

$$\mathbf{W}_{LS} = \mathbf{I}_{r \times r}$$

$$\mathbf{W}_{MV} = \text{diag} \left(\left(1 - \frac{\sigma_{\text{noise}}^2}{\lambda_1^2} \right), \dots, \left(1 - \frac{\sigma_{\text{noise}}^2}{\lambda_r^2} \right) \right). \quad (4)$$

It should be noted that, \mathbf{U}_1 and \mathbf{V}_1 , of LS and MV estimates are the same, but the singular values are different.

Fig. 1 shows the signal extracted by SSA-MV in [28] along with fitted trend obtained by Synthesis Diffusion Degradation (SDD) model as the most widely accepted model in analysing *Bcd* gradient. As it appears SSA-MV yields more promising result for *Bcd* signal extraction.

2.2. SSA combined with AR model

Among many applications of the microarray technique, the study of rhythmic cellular processes has been considered as an important application. Rhythmic cellular processes are mainly regulated by different gene products, and can be measured by using multiple DNA microarray experiments. Having a group of gene experiments over a time period, a time series of gene expression related to the rhythmic behaviour of that specific gene will be achieved. Several studies on extracting the regulatory information from time-series microarray data and detection of cyclic and non-uniformly sampled short time series of gene expressions can be found in [30–32]. The characteristics of the rhythmic gene expression is as follows [33]: the number of time points and cycles related to a profile is usually very few. For example, the 14 time point elutriation observations may be in constitution of just one cell-cycle [20], this data set usually contains many missing values which need to be determined [34], the intervals spaced between time points are not equal and the gene expression data is extremely noisy [33].

In 2008, Du et al. implied SSA for analysing microarray results for extracting the dominant trend from noisy expression profiles and reducing the effect of noise [12]. The authors have also proposed a new procedure for analysing the periodicity of the transcriptome of the intraerythrocytic developmental cycle (IDC) by combining autoregressive (AR) model and SSA. This combination enabled them to successfully identify almost 90% of genes (4496 periodic profiles) in *P. falciparum*, which was a considerable achievement in terms of detecting 777 additional periodic genes in comparison to the results in [35]. Fig. 2 depicts the improvement yielded by using SSA in analysing periodicity. As shown in this figure, due to the

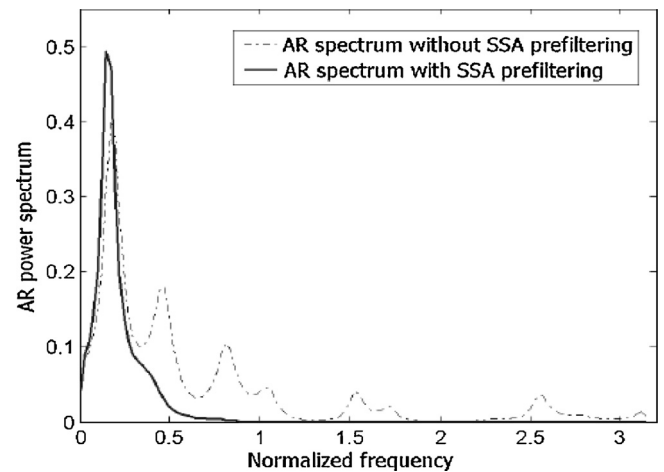


Fig. 2. The AR spectra of the expression profile of Dihydrofolate Reductase-Thymidylate Synthase with and without SSA filtering [12].

removal of noise using the SSA, spurious peaks are eliminated from the spectrum.

Four subsequent research works followed this procedure and successfully improved the capability of detecting periodicity from 60% to 80% [36–39].

According to [33], the periodic profiles can be detected by combining SSA and AR as follows:

- At the first step SSA is used for the reconstruction of each expression data. For this aim only those expression profiles with sum of first two eigenvalues over the sum of all eigenvalues greater than 0.6 are selected for reconstruction.
- The second step is devoted to the calculation of the AR spectrum, frequency f_i at peak value point and the ratio of the power in f_i Regions of Interest (ROI) of the reconstructed profiles achieved in the first step. It should be noted that to obtain the ratio between the power of the signal within the frequency band $[f_{i-1}, f_{i+1}]$ and the total power we follow $S = \text{power}_i / \text{power}_{\text{total}}$.
- If the obtained power ratio S is larger than 0.7, the corresponding profile would be classified as periodic [33].

Presented below is a summary of the theory combining SSA and AR based on [39]:

Let the microarray data be $g \times N$ matrix ($g \gg N$), where g is the number of gene expression profiles and N corresponds to the

number of samples. Time series gene expression $Y_N = (y_1, \dots, y_N)$ can then be written in a form of an $AR_{(p)}$ model by forward-backward linear prediction as follows [39]:

$$y_i = \alpha_1 y_{i-1} + \alpha_2 y_{i-2} + \dots + \alpha_p y_{i-p} \quad (i = p + 1, \dots, N). \quad (5)$$

Using the forward prediction linear system the AR coefficients $(\alpha_1, \alpha_2, \dots, \alpha_p)$ can be estimated and gene expression can be recognised as a linear system. In an $AR_{(p)}$ model, p has to be set in a way that the system models the desired trend and avoids redundant data such as noise. By choosing m gene expression profiles, the number of linear equations is equal to $m \times 2(N - p)$.

As mentioned above, the SSA technique is often applied prior to spectral analysis, as a filtering method, in order to achieve a better accuracy. This is done by ignoring the small singular values in the reconstruction stage. Removing the noise component from the original noisy signal Y_N yielding the noise reduced series \hat{Y}_N which can consequently be used for estimating the AR coefficients α_i ($i = 1, \dots, p$), and ultimately estimated α_i are obtained using Yule-Walker equations [39].

2.3. SVD

Although SVD is an stage of SSA technique, it has also been used independently as a very useful and applicable tool for analysing the microarrays data (see for example, [40]) in which the challenging problem of eliminating cross hybridization in real-time microarray data was considered. Vikalo et al. in [40] evaluated multiple techniques such as SVD for separating the components of the compound signal, to find a better estimation of the amounts of both hybridising and cross-hybridising targets.

However, the microarray technology has become increasingly affordable, the data is still very complicated to analyse. In 2010 Rau et al. in [41] presented an algorithm to infer the structure of gene regulatory networks using an empirical Bayes estimation procedure for the hyperparameters of a linear feedback state-space model and used SVD of the block-Hankel matrix for model selection. This approach enabled them to significantly reduce the computation time required for running the algorithm, as it eliminates the need to run the algorithm over a wide range of values for the hidden state dimension. Moreover, as discussed in [42] the numerical computations in SVD even for a huge data set such as microarray data is not time extensive. In that study a new approach for ranking genes is presented which is based on the genes degree of regulation and the authors showed that the ranking of genes according to regulation is genetically more meaningful rather than using the absolute expression or variation over time and this method can be a valuable aid in finding the regulatory pathways and networks. In that work SVD was performed to the block-Hankel matrix of observation autocovariances estimated from the gene expression data, and the number of singular values related to the large magnitude was performed as an estimate for the best state space dimension. The singular values of the estimated Hankel autocovariance matrix were calculated and standardised to a 0–1 scale. The number of singular values of magnitude larger than the threshold was used as an estimate for the state space dimension.

Based on the mentioned studies, SSA procedure is a flexible technique for the signal extraction and gene expression degradation modelling. The feasibility of capturing the signal from segmentation genes profile in *Drosophila* embryos and studying the rhythmic behaviour of a specific gene suggest that this technique may be of general use in evaluating other expressional systems. Thus, this method can be a valuable aid in analysing spatially inhomogeneous noisy data.

3. Application of two dimensional SSA

The 2D-SSA approach has been used to process two-dimensional scalar fields [8]. The first difference between 2D-SSA and SSA is about the window length. In 2D-SSA analysis we need to choose two different values for the window length L (L_1, L_2), whilst in univariate SSA we only require to select one window length. Note that if $L_2 = 1$, then 2D-SSA is equivalent to SSA. By choosing $L_2 = M$, the interaction among different series is taken into account [43]. For more information see [44].

It is worthy to mention that the gene expression can be traced either in just anterior–posterior (AP) axis or both AP and dorso-ventral (DV) axis which the latter case is the subject to 2D-SSA. The data points used in 2D-SSA study attributes to the intensity levels for the positions along both AP and DV axis and are considered as a sequenced series.

Bcd is a transcriptional regulator of downstream segmentation genes where the alteration in *Bcd* gradient shifts the downstream patterns [45]. However, it has been accepted that in the embryos, zygotic gene products are considerably more precisely positioned than the gradients related to maternal genes, which indicates an embryonic error depletion process [46]. In 2011 the anterior–posterior (AP) segmentation of *Drosophila* was studied in [47] to determine how gene regulation dynamics control noise. In this research the activation of the *hb* gene by the *Bcd* protein gradient in the anterior part was studied by modelling the noise observed in *hb* regulation using a chemical master equation approach. For solving this model, the MesoRD software package has been used which mainly follows a stochastic approach [48] and the results indicate that *Hb* output noise is mostly dependent on the transcription and translation dynamics of its own expression, and that the multiple *Bcd* binding sites located in the *hb* promoter also improve pattern formation [49].

Noise in that study is calculated by:

$$\sqrt{\frac{\sum [(data - trend)/trend]^2}{m - 1}}, \quad (6)$$

where data is background-removed intensity of an energid (nucleus plus surrounding cytoplasm), trend is the signal extracted using 2D SSA (AP and DV) and m indicates the positions. This measurement is obtained for the activated region (15–45% EL).

In 2012 Golyandina et al. used 2D SSA to measure between-nucleus variability (noise) seen in the gradient of *Bcd* in *Drosophila* embryos [50]. In that paper, they measured the noise using 2D SSA for comparing the results of fixed immunostained embryos with live embryos with fluorescent *bcd* (*bcd*-GFP).

It should be noted that using SSA for signal extraction gives the ability of using both dimensions (AP and DV) which leads to a more reliable result as in this case more detailed information regarding the data has been considered by the technique to give the results.

4. Conclusion

The aim of this paper was devoted to review the applications of SSA in genetics studies. Previous research has shown that the SSA technique is very effective in signal extraction and noise filtering in genetics data.

Theoretical developments presented here as two-dimensional SSA and SSA based on minimum variance have also enabled the researchers to achieve enhanced results and a better estimation of extracted signal. As a non-parametric method SSA has given us very promising results which are more reliable than those obtained by other methods. However, SSA has not revealed its full potential yet, areas like optimising the embedding dimension and change point detection are still open to pursue.

References

- [1] Allison DB, Cui X, Page GP, Sabripour M. Microarray data analysis: from disarray to consolidation and consensus. *Nat Rev Genet* 2006;7(1):55–65.
- [2] Goeman JJ, Bhlmann P. Analyzing gene expression data in terms of gene sets: methodological issues. *Bioinformatics* 2007;23(8):980–7.
- [3] VanGuilder HD, Vrana KE, Freeman WM. Twenty-five years of quantitative PCR for gene expression analysis. *Biotechniques* 2008;44(5):619.
- [4] Bustin SA. The reproducibility of biomedical research: sleepers awake! *BDQ* 2014;2:35–42.
- [5] Huggett J, O'Grady J, Bustin S. How to make mathematics biology's next and better microscope. *BDQ* 2014;1:A1–3.
- [6] Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am J Hum Genet* 1996;58(6):1347.
- [7] Grimm O, Coppey M, Wieschaus E. Modelling the bicoid gradient. *Development* 2010;137(14):2253–64.
- [8] Golyandina N, Nekrutkin V, Zhigljavsky AA. Analysis of time series structure: SSA and related techniques. CRC Press; 2010.
- [9] Hassani H. Singular spectrum analysis: methodology and comparison. *JDS* 2007;5:239–57.
- [10] Sanei S, Ghodsi M, Hassani H. An adaptive singular spectrum analysis approach to murmur detection from heart sounds. *IPEM* 2011;33(3):362–7.
- [11] Xie HB, Guo T, Sivakumar B, Liew AWC, Dokos S. Symplectic geometry spectrum analysis of nonlinear time series. *Proc R Soc A* 2014;470(2170):20140409.
- [12] Du LP, Wu SH, Liew AWC, Smith DK, Yan H. Spectral analysis of microarray gene expression time series data of *Plasmodium falciparum*. *IJBRA* 2008;4(3):337–49.
- [13] Tang VT, Yan H. Noise reduction in microarray gene expression data based on spectral analysis. *IJMLC* 2012;3(1):51–7.
- [14] Broomhead DS, King GP. Extracting qualitative dynamics from experimental data. *Physica D* 1986;20:217–36.
- [15] Vautard R, Yiou P, Ghil M. Singular-spectrum analysis: a toolkit for short, noisy chaotic signal. *Physica D* 1992;58:95–126.
- [16] Ghil M, Taricco C. Advanced spectral analysis methods. In: Castagnoli GC, Provenzale A, editors. Past and present variability of the solar-terrestrial system: measurement, data analysis and theoretical model. IOS Press; 1997. p. 137–59.
- [17] Elsner JB, Tsonis AA. Singular spectral analysis. A new tool in time series analysis. Plenum Press; 1996.
- [18] Danilov D. Principal components in time series forecast. *J Comp Graph Stat* 1997;6:112–21.
- [19] Holloway DM, Harrison LG, Kosman D, Vanario Alonso CE, Spirov AV. Analysis of pattern precision shows that *Drosophila* segmentation develops substantial independence from gradients of maternal gene products. *Dev Dyn* 2006;235(11):2949–60.
- [20] Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, et al. Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Mol Biol Cell* 1998;9(12):3273–97.
- [21] Alexandrov T, Golyandina N, Spirov A. Singular spectrum analysis of gene expression profiles of early *Drosophila* embryo: exponential-in-distance patterns. *Res Lett Signal Process* 2008;2008:12.
- [22] Gregor T, Wieschaus EF, McGregor AP, Bialek W, Tank DW. Stability and nuclear dynamics of the bicoid morphogen gradient. *Cell* 2007;130(1):141–52.
- [23] Hilfinger A, Paulsson J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *PNAS* 2011;108(29):12167–72.
- [24] Klebanov L, Yakovlev A. How high is the level of technical noise in microarray data. *Biol Direct* 2007;2(9).
- [25] <http://cran.r-project.org/web/packages/Rssa/Rssa.pdf>
- [26] Surkova S, Kosman D, Koslov K, Manu E, Myasnikova A, Samsonova A, et al. Characterization of the *Drosophila* segment determination morphome. *Dev Biol* 2007;313(2008):844–62.
- [27] Lopes FJP, Vieira FMC, Holloway DM, Bisch PM, Spirov AV. Spatial bistability generates hunchback expression sharpness in the *Drosophila* embryo. *PLoS Comput Biol* 2008;4(9):e1000184.
- [28] Hassani H, Ghodsi Z. Pattern recognition of gene expression with singular spectrum analysis. *Med Sci* 2014;2(3):127–39.
- [29] Hassani H. Singular spectrum analysis based on the minimum variance estimator. *Nonlinear Anal: Real World Appl* 2010;11(3):2065–77.
- [30] Liew AWC, Law NF, Cao XQ, Yan H. Statistical power of Fisher test for the detection of short periodic gene expression profiles. *Pattern Recogn* 2009;42(4):549–56.
- [31] Liew AWC, Xian J, Wu S, Smith D, Yan H. Spectral estimation in unevenly sampled space of periodically expressed microarray time series data. *BMC Bioinform* 2007;8(1):137.
- [32] Yeung LK, Szeto LK, Liew AWC, Yan H. Dominant spectral component analysis for transcriptional regulations using microarray time-series data. *Bioinformatics* 2004;20(5):742–9.
- [33] Liew AWC, Yan H. Reliable detection of short periodic gene expression time series profiles in DNA microarray data. *International conference SMC* 2009;427:4–427, 9.
- [34] Gan X, Liew AWC, Ya H. Microarray missing data imputation based on a set theoretic framework and biological consideration. *Nucleic Acids Res* 2006;34(5):1608–19.
- [35] Bozdech Z, Llinas M, Pulliam BL, Wong ED, Zhu JC, DeRisi JL. The transcriptome of the intraerythrocytic developmental cycle of *Plasmodium falciparum*. *PLoS Biol* 2003;1(1):1–16.
- [36] Tang TY, Yan H. Identifying periodicity of microarray gene expression profiles by autoregressive modeling and spectral estimation. Ninth international conference on machine learning and cybernetics 2010;306:2–3066.
- [37] Tang TY, Liew AWC, Yan H. Periodicity analysis of DNA microarray gene expression time series profiles in mouse segmentation clock data. *Stat Interface* 2010;3(3):413–8.
- [38] Tang TY, Liew AWC, Yan H. Analysis of mouse periodic gene expression data based on singular value decomposition and autoregressive modeling. *IAENG*; 2010.
- [39] Tang TY, Yan H. Noise reduction in microarray gene expression data based on spectral analysis. *Int J Mac Learn Cybern* 2012;3(1):51–7.
- [40] Vikalo H, Hassibi B, Hassibi A. Modeling and estimation for real-time microarrays. *IEEE J Sel Topics Signal Process* 2008;2(3):286–96.
- [41] Rau A, Jaffrezic F, Foulley JL, Doerge RW. An empirical Bayesian method for estimating biological networks from temporal microarray data. *Stat Appl Genet Mol Biol* 2010;9(1):1544–6115.
- [42] Bremer M, Doerge RW. The KM-algorithm identifies regulated genes in time series expression data. *Adv Bioinform* 2009;115:5–1165.
- [43] Zhang J, Hassani H, Xie H, Wang S. Estimating multi-country prosperity index: a two-dimensional singular spectrum analysis approach. *J Syst Sci Complex* 2013;20:1–20.
- [44] Golyandina NE, Usevich KD. 2D-extension of singular spectrum analysis: algorithm and elements of theory. *Matrix methods: theory, algorithms. Applications* World Scientific; 2010.
- [45] Porcher A, Dostatni N. The bicoid morphogen system. *Curr Biol* 2010;20(5):249–54.
- [46] Spirov AV, Holloway DM. Making the body plan: precision in the genetic hierarchy of *Drosophila* embryo segmentation. *In Silico Biol* 2003;3(1):89–100.
- [47] Holloway DM, Lopes FJ, da Fontoura Costa L, Travençolo BA, Golyandina N, Usevich K, et al. Gene expression noise in spatial patterning: hunchback promoter structure affects noise amplitude and distribution in *Drosophila* segmentation. *PLoS Comput Biol* 2011;7(2):e1001069.
- [48] <http://mesord.sourceforge.net>
- [49] Holloway DM, Spirov AV. Gene expression noise in embryonic spatial patterning. In: 21st international conference on noise and fluctuations. 2011.
- [50] Golyandina NE, Holloway DM, Lopes FJP. Measuring gene expression noise in early *Drosophila* embryos: nucleus-to-nucleus variability. *Proc Comput Sci* 2012;9:373–82.