

Article

Pattern Recognition of Gene Expression with Singular Spectrum Analysis

Hossein Hassani * and Zara Ghodsi

The Staistical Research Centre, Bournemouth University, Bournemouth BH8 8EB, UK;
E-Mail: zghodsi@bournemouth.ac.uk

* Author to whom correspondence should be addressed; E-Mail: hhassani@bournemouth.ac.uk;
Tel.: +44-1202978609.

Received: 13 April 2014; in revised form: 19 June 2014 / Accepted: 24 June 2014 /

Published: 1 July 2014

Abstract: *Drosophila* segmentation as a model organism is one of the most highly studied. Among many maternal segmentation coordinate genes, bicoid protein pattern plays a significant role during *Drosophila* embryogenesis, since this gradient determines most aspects of head and thorax development. Despite the fact that several models have been proposed to describe the bicoid gradient, due to its association with considerable error, each can only partially explain bicoid characteristics. In this paper, a modified version of singular spectrum analysis is examined for filtering and extracting the bicoid gene expression signal. The results with strong evidence indicate that the proposed technique is able to remove noise more effectively and can be considered as a promising method for filtering gene expression measurements for other applications.

Keywords: *Drosophila melanogaster*; segmentation; signal extraction; singular spectrum analysis

1. Introduction

Segmentation is a crucial stage of many animal body plans. One of the best understood examples of segmentation is the patterning along the future head to tail axis of the *Drosophila melanogaster*. Since most of the developmental regulators identified in *Drosophila* are conserved genes, the *Drosophila* model has a profound influence on other fields, including medicine [1]. During *Drosophila* embryogenesis the

segmented body plan is established through a cascade of both maternally and zygotically expressed segmentation genes [2].

Specifically, in the modern formulation, cells or nuclei would respond to different concentration thresholds of the morphogen by the expression of specific target genes, thereby translating the differential morphogen concentration into patterns of gene expression [3]. The zygotic segmentation genes were first identified in a large scale screen for embryonic lethal mutations that affect the body pattern [4]. Bicoid, which encodes a homeobox transcription factor, was the first gene to be identified that fulfils the criteria of such a morphogen [5].

Of the many maternal coordinate genes, here we are concerned only with those in the anterior groups that regulates zygotic segmentation genes, namely bicoid (*bcd*) (the italic lower case *bcd* represents the gene and *Bcd* refers to protein). In *Drosophila melanogaster*, *bcd* mRNA, transcribed in the mother, is deposited at the anterior end of the fertilized egg, and the *Bcd* protein is translated and diffuses to form an anteroposterior (AP) concentration gradient [6]. This gradient determines most aspects of head and thorax development. In *Bcd*-deficient embryos, the thoracic segments are developed from more anterior regions than normal, and less of the body is devoted to the head [1].

Once the *Bcd* gradient is established, target genes, such as the hunchback and Krüppel gap genes, are transcribed at different concentrations of *Bcd* and thus at different positions in the embryo [7]. Several models have been proposed to describe the *Bcd* gradient in the *Drosophila* embryo since its discovery in 1988. However, none of those are fully consistent with all available experimental observations; each can only partially account for and explain experimental findings on the *Bcd* gradient [8]. Among all extensive studies, the most widely accepted model was formulated before the identification of *Bcd* [5], which is called the simple diffusion model [3] (also referred to as the *SDD* model). The *SDD* model has been widely used to explain the exponential gradient of *Bcd* [9]:

$$B = Ae^{-x/\lambda} \quad (1)$$

where A is the amplitude, x is distance from the anterior, and λ is the length parameter [10]. It should be noted that λ is obtained by fitting an exponential model to the *Bcd* intensity profile, and computing the position at which the concentration has dropped to $1/exp$ of the maximal value at the anterior (at $x = 0$) [8].

However, one of the challenging issues in analysing *Bcd* is its association with both intrinsic and extrinsic noise, which limits the performance of the models and techniques, and consequently affects the accuracy of results [11,12]. Therefore, employing an effective method for dealing with noisy *Bcd* is essential.

There are two main approaches for fitting a model to a noisy *Bcd*. The first approach ignores the presence of noise and directly fits a model to noisy *Bcd* data (like the *SDD* model). The second approach, which is often more effective, seeks to reduce the *Bcd* noise level via filtering methods, and then a model is fitted to the filtered data [13]. It should be noted that the selection of an appropriate filtering method is crucial for the effectiveness of the second approach. For the standard filtering methods, assuming stationarity for the data, linearity of the model and normality of the residuals can provide only an approximation of the true situation [18]. Therefore, a method that does not depend on these assumptions could be very useful for modelling and filtering *Bcd* data. Singular spectrum analysis (SSA) is a very

good example as it is nonparametric and therefore does not require any assumptions [14] and basic SSA has been applied since 2006 for extracting the *Bcd* signal from its noisy datasets; for example see [2,6,19–21,23]. Accordingly, we introduce a new method of SSA for filtering and signal extraction of *Bcd*, which is a powerful technique for filtering noisy signal in time series analysis and signal processing.

The basic SSA method consists of two complementary stages: decomposition and reconstruction. Each stage includes two separate steps. At the first stage the *Bcd* is decomposed into the sum of a small number of independent and interpretable components, such as a slowly varying trend and a structureless noise [16]. At the second stage, the noise-free *Bcd* is reconstructed [17].

Here, the SSA technique based on minimum variance is utilized as it has been shown to be more effective than the basic version of it [22]. This enables extraction of the *Bcd* signal more accurately in comparison with the basic version [2,6,23].

In recent years the SSA technique has been developed and applied to many biomedical and genetic data (see, for example, [24,25]). It has also been used for filtering microarray gene expression data [15,26], curve fitting in growth curve models [13], and detecting the gene expression pattern in [2,6,23]. It has also been used for reliably imaging radioactive seeds [27].

The rest of the paper is organised as follows: the SSA technique is briefly described in Section 2. The empirical results are represented in Section 3, and finally Section 4 is devoted to summary and conclusions.

2. Singular Spectrum Analysis (SSA)

2.1. LS and MV Estimators

A short description of the SSA based on minimum variance technique is given below, and in doing so we mainly follow [22] where a more detailed description is made available.

Consider a noisy signal vector $Y_N = (y_1, \dots, y_N)^T$ of length N .

$$Y_N = S_N + E_N \tag{2}$$

where S_N represents the signal component and E_N the noise component. Let $K = N - L + 1$, where L is some integer called the window length (we can assume $L \leq N/2$). Define the so-called “trajectory matrix”:

$$\mathbf{X} = (x_{ij})_{i,j=1}^{L,K} = \begin{pmatrix} y_1 & y_2 & y_3 & \dots & y_K \\ y_2 & y_3 & y_4 & \dots & y_{K+1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ y_L & y_{L+1} & y_{L+2} & \dots & y_N \end{pmatrix} \tag{3}$$

The columns X_j of \mathbf{X} , considered as vectors, lie in an L -dimensional space \mathbb{R}^L . It is obvious that:

$$\mathbf{X} = \mathbf{S} + \mathbf{E} \tag{4}$$

where \mathbf{S} and \mathbf{E} represent Hankel matrices of the signal S_N and noise E_N , respectively. The singular value decomposition (SVD) of the trajectory matrix \mathbf{X} can be written as:

$$\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T \tag{5}$$

where $\mathbf{U} \in \mathbb{R}^{L \times K}$ is the matrix consists of the normalized eigenvector U_i corresponding to the eigenvalue λ_i ($i = 1, \dots, L$), $\mathbf{V} \in \mathbb{R}^{K \times K}$, is the matrix contains the principal components defined as $V_i = \mathbf{X}^T U_i / \sqrt{\lambda_i}$, and $\mathbf{\Sigma} = \text{diag}(\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_L)$. The diagonal elements of $\mathbf{\Sigma}$ are called singular value of \mathbf{X} , and their set is called the singular value spectrum.

The signal subspace methods are based on the assumption that the vector space of the noisy time series (signal) can be split in mutually orthogonal noise and “signal+noise” subspaces. Thus, by adapting the weights of the different singular components, an estimate of the Hankel matrix \mathbf{X} , which corresponds to noise reduced series, can be achieved:

$$\mathbf{X} = \mathbf{U}(\mathbf{W}\mathbf{\Sigma})\mathbf{V}^T \tag{6}$$

where \mathbf{W} is the diagonal matrix containing the weights. Now, the problem is choosing the weight matrix \mathbf{W} . Next we consider the problem of choosing this matrix using different criteria. The SVD of the matrix \mathbf{X} can be written as:

$$\mathbf{X} = [\mathbf{U}_1 \ \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & \mathbf{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} \tag{7}$$

where $\mathbf{U}_1 \in \mathbb{R}^{L \times r}$, $\mathbf{\Sigma}_1 \in \mathbb{R}^{r \times r}$ and $\mathbf{V}_1 \in \mathbb{R}^{K \times r}$. We can also represent the SVD of the Hankel matrix of the signal s_T as:

$$\mathbf{S} = [\mathbf{U}_{1s} \ \mathbf{U}_{2s}] \begin{bmatrix} \mathbf{\Sigma}_{1s} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_{1s}^T \\ \mathbf{V}_{2s}^T \end{bmatrix} \tag{8}$$

It is clear that the Hankel matrix \mathbf{S} cannot be reconstructed exactly if it is perturbed by noise.

2.2. LS Estimate of S

Let us consider the assumption that the matrix $\mathbf{X}_{L \times K}$ is rank deficient, *i.e.*, rank $\mathbf{X} = r$ and $r < L < K$. The simplest estimate of \mathbf{S} is obtained when we approximate \mathbf{S} by a matrix of rank r in the LS sense:

$$\min \|\mathbf{X} - \hat{\mathbf{S}}_{LS}\|_F^2 \tag{9}$$

where $\|\cdot\|_F$ is the *Frobenius* norm. That is, the LS estimate is obtained by setting the smallest singular value to zero ($\lambda_{r+1} = 0, \dots, \lambda_L = 0$) in Equation (7):

$$\hat{\mathbf{S}}_{LS} = [\mathbf{U}_1 \ \mathbf{U}_2] \begin{bmatrix} \mathbf{\Sigma}_1 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{V}_1^T \\ \mathbf{V}_2^T \end{bmatrix} = \mathbf{U}_1 \mathbf{\Sigma}_1 \mathbf{V}_1^T \tag{10}$$

The \mathbf{S}_{LS} estimate removes the noise subspace but keeps the noisy signal uncorrelated in the “signal+noise” subspace. The disadvantage of LS is that the performance of the LS estimator is crucially dependent on the estimation of the signal rank r . That is, selecting singular values in LS is a binary approach. The main advantage of the LS estimate is that one does not need to consider any assumptions either about the signal or the noise.

2.3. MV Estimate of S

The aims of the noise reduction can be considered as follows: (1) separate the “signal+noise” subspaces from the noise-only subspace; (2) remove the noise-only subspace; (3) ideally, remove the noise components in the “signal+noise” subspaces. The first two steps can be achieved by the least squares estimate, while the MV estimate allows us to achieve the third aim as well. However, one should consider some assumptions to obtain the MV estimate [22] and if the assumptions are met, one can obtain the MV estimate as follows [28,31]. Given the matrix \mathbf{X} , with $\text{rank } \mathbf{X} = \text{rank } \mathbf{N} = L$ and also $\text{rank } \mathbf{S} = r$. Find the matrix $\mathbf{P} \in \mathbb{R}^{K \times K}$ that minimises $\min \|\mathbf{XP} - \mathbf{S}\|_F^2$ and the solution is obtained by $\mathbf{P} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}$. Therefore, the MV estimate of \mathbf{S} is $\mathbf{XP} = \mathbf{X}(\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{S}$. Using the SVD of the \mathbf{X} , we can obtain $\mathbf{XP} = \mathbf{U}\mathbf{U}^T \mathbf{S}$.

That is, the MV estimate of \mathbf{S} can be interpreted as an orthogonal projection of \mathbf{S} onto the column space of \mathbf{X} because $\mathbf{U}\mathbf{U}^T$ is the associated projection matrix. Note also that $\text{rank}(\mathbf{XP}) = \text{rank}(\mathbf{S}) = r$. Let us now consider an alternative form of the SVD of the matrix \mathbf{X} using the SVD of \mathbf{S} Equation (8) as follows:

$$\begin{aligned} \mathbf{X} &= \mathbf{S} + \mathbf{N} = \mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} \mathbf{V}_{1s}^T + \mathbf{N}\mathbf{V}_{1s} \mathbf{V}_{1s}^T + \mathbf{N}\mathbf{V}_{2s} \mathbf{V}_{2s}^T \\ &= [(\mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})(\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{-1/2} \quad \sigma_{noise}^{-1} \mathbf{N}\mathbf{V}_{2s}] \\ &\times \begin{bmatrix} (\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{1/2} & \mathbf{0} \\ \mathbf{0} & \sigma_{noise} \mathbf{I} \end{bmatrix} \begin{bmatrix} \mathbf{V}_{1s}^T \\ \mathbf{V}_{2s}^T \end{bmatrix} \end{aligned} \tag{11}$$

As it appears from Equation (11), the middle matrix is diagonal, and the left and right matrices have orthonormal columns. Therefore, Equation (11) can be considered as an alternative form of the SVD of \mathbf{X} , and the singular values of \mathbf{X} are:

$$\begin{aligned} \boldsymbol{\Sigma}_1 &= (\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{1/2} \\ \boldsymbol{\Sigma}_2 &= \sigma_{noise} \mathbf{I} \end{aligned} \tag{12}$$

Hence, the singular values in $\boldsymbol{\Sigma}_2$ can be interpreted as a noise threshold, which permits estimating σ_{noise} from $\boldsymbol{\Sigma}_2$ in Equation (12). We can also consider the following submatrices:

$$\begin{aligned} \mathbf{U}_1 &= (\mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s})(\boldsymbol{\Sigma}_{1s}^2 + \sigma_{noise}^2 \mathbf{I})^{-1/2} \\ &= (\mathbf{U}_{1s} \boldsymbol{\Sigma}_{1s} + \mathbf{N}\mathbf{V}_{1s}) \boldsymbol{\Sigma}_1^{-1} \\ \mathbf{U}_2 &= \sigma_{noise}^{-1} \mathbf{N}\mathbf{V}_{2s} \\ \mathbf{V}_1 &= \mathbf{V}_{1s} \\ \mathbf{V}_2 &= \mathbf{V}_{2s} \end{aligned} \tag{13}$$

Now, the MV estimate of \mathbf{S} :

$$\mathbf{U}_1 \boldsymbol{\Sigma}_1^{-1} (\boldsymbol{\Sigma}_1^2 - \sigma_{noise}^2 \mathbf{I}) \mathbf{V}_1^T \tag{14}$$

2.4. Weight Matrix W

Let us consider again the weight matrix \mathbf{W} based on the LS and MV estimates. As it is appears from Equations (10) and (14), the left and right singular vectors, \mathbf{U}_1 and \mathbf{V}_1 , of LS and MV estimates are the

same, but the singular values are different. The LS and MV estimates can be defined based on the weight matrix $\mathbf{W}_{r \times r}$ as follows:

$$\begin{aligned} \hat{\mathbf{S}}_{LS} &= \mathbf{U}_1(\mathbf{W}_{LS}\Sigma_1)\mathbf{V}_1^T \\ \hat{\mathbf{S}}_{MV} &= \mathbf{U}_1(\mathbf{W}_{MV}\Sigma_1)\mathbf{V}_1^T \end{aligned} \tag{15}$$

where

$$\begin{aligned} \mathbf{W}_{LS} &= \mathbf{I}_{r \times r} \\ \mathbf{W}_{MV} &= \text{diag} \left(\left(1 - \frac{\sigma_{noise}^2}{\lambda_1^2}\right), \dots, \left(1 - \frac{\sigma_{noise}^2}{\lambda_r^2}\right) \right) \end{aligned} \tag{16}$$

Here the capability of the SSA technique based on the minimum variance estimator, in reconstructing, is tested by applying the technique to the simple exponential series:

$$B = Ae^{-x/\lambda} + \epsilon_n \tag{17}$$

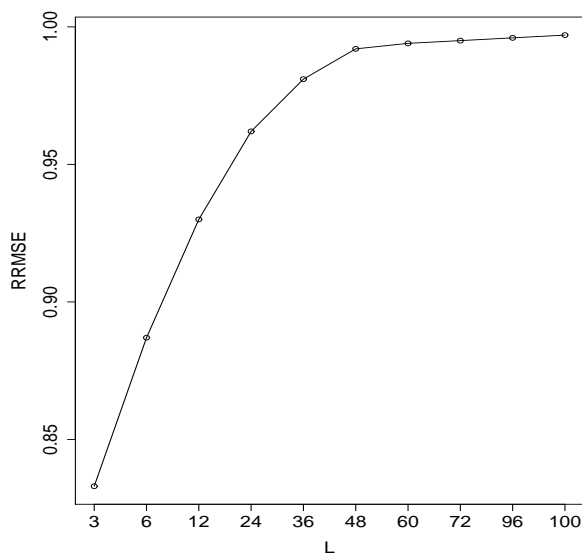
where ϵ_n is the not normally distributed noise series. In total 200 observations were generated and we added different level of noise to each point of the original series. The simulation was repeated 1000 times. Different values of window length L were considered to examine the sensitivity of the SSA technique for different L . Based on the theoretical results L should be large enough but not greater than $N/2$ [29]. Here, for reconstruction the first two eigenvalues are selected and the ratio of root mean square error (RRMSE) is used to compare two noise reduction methods [30]:

$$\text{RRMSE} = \frac{\text{RMSE}(SSA_{MV})}{\text{RMSE}(SSA)} = \frac{\left(\sum_{i=1}^N (s_i - \hat{s}_i)^2\right)^{1/2}}{\left(\sum_{i=1}^N (s_i - \tilde{s}_i)^2\right)^{1/2}}$$

where, \hat{s}_i are the estimated values of s_i , obtained with SSA_{MV} and \tilde{s}_i are the estimated values of s_i obtained by SSA_{LS} and N is the series length. If $\text{RRMSE} < 1$, then the SSA_{MV} procedure outperforms the alternative method. Conversely, $\text{RRMSE} > 1$ would indicate that the performance of the corresponding SSA_{MV} procedure is worse than the competing method.

Figure 1 shows the RRMSE of SSA_{MV}/SSA_{LS} . As it appears, the SSA_{MV} has better performance for small window length and the RRMSE tends to 1 as the window length increases, confirming that both methods have similar performance for large window length. Also it can be seen that there is a gradual increase in RRMSE with window length. For a window length of 3, the performance of SSA_{MV} is up to 14% better than SSA_{LS} in reconstruction. However, there is no significant discrepancy between the performance of SSA_{MV} and SSA_{MV} for window length greater than 36. For more detailed information see [22].

Figure 1. The RRMSE of SSA_{MV}/SSA_{LS} in the reconstruction of noisy exponential series.



3. Empirical Results

3.1. Simulated Series

In this paper, a series of simulated data are used to evaluate the performance of two different approaches for extracting *Bcd* signal from its noisy datasets. The first part assesses the SDD model as the most common model in studying *Bcd*, which follows an exponential curve. In this approach the performance of SDD is evaluated before and after filtering. The second part studies the efficiency of using SSA_{MV} directly for trend extraction.

For starting the simulation, an exponential curve drawn from the SDD model was considered as the benchmark, then random errors ϵ of a normal distribution with zero mean and variance σ_ϵ^2 were added to this curve to obtain a noisy datasets. This simulation was repeated 10,000 times. Finally, by fitting the nonlinear model to the noisy simulated *Bcd* series, parameters A and λ were estimated.

Table 1 presents the values of estimated parameters and the related standard deviation. As it appears, following the application of SSA, these values are more robust and accurate than the values prior to filtering.

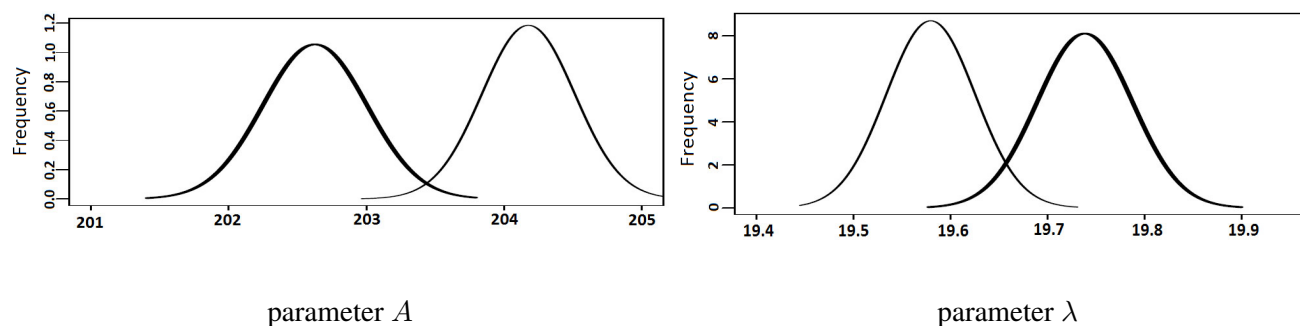
Table 1. Standard deviation of parameters before and after applying SSA.

Parameter	Original Value	Average		S.D		Ratio
		before	after	before	after	
A	200	204.2	202.6	4.19	2.64	$\frac{2.64}{4.19} = 0.63$
λ	20	19.57	19.76	0.41	0.26	$\frac{0.26}{0.41} = 0.62$

Figure 2 also shows the normal distribution of the estimated parameters A and λ before and after filtering. As it appears, parameters show different values in noisy and noise-free series, which confirms

that noise reduction does help to obtain a better model. Thus, if the user is inclined to apply a parametric model such as SDD, it is strongly suggested that they first reduce the noise with any noise reduction method (here for example SSA) and then fit a model to the noise reduced Bcd , or they can just apply SSA, which is a powerful nonparametric method for both noise reduction and trend extraction.

Figure 2. Distribution of the estimated parameters of A and λ for noisy Bcd and noise-reduced Bcd (thick line).



For the SSA analysis, since the length of each series is 100 (one observation per egg length percentage), the window length of 50 was chosen for trend extraction and reconstruction. Choosing $L = 50$ and performing the SVD step of the trajectory matrix \mathbf{X} , 50 eigentriples were obtained, ordered by their contribution (share) in the decomposition stage. We shall say that the series Y_N is not complex if Y_N is well approximated by a series with small rank d . For example, series $y_i = e^{\alpha i}$ ($i = 1, \dots, N$) has rank 1. For all $2 \leq L \leq N - 1$, $y_i = by_{i-1}$ where $b = e^\alpha$. It should be noted that the number of eigentriples selected as corresponding to the series S_N has to be at least d . For example, if $y_i = e^\alpha + \epsilon_i$ then the window length L should be at least 2 and the first eigentriple is enough for reconstructing the original series if $\|S_N\| \gg \|E_N\|$. Accordingly, the first eigentriple is selected for filtering and trend extraction.

Furthermore, the ratio of Root Mean Square Error (RRMSE) is used to measure the difference between the estimated values (fitted by SDD model and reconstructed by SSA) and the actual values.

Table 2 shows the results. As it appears from Table 2, a significant reduction in the RMSE value is achieved by SSA, confirming that these results are more accurate than the estimated points by SDD model. For all simulation runs (here 5 times) SSA outperforms the SDD model ranging from 42% to 46%. For instance, for the case 1, first row in Table 2, the extracted signal by SSA is 45% more accurate than the SDD model.

Table 2. The RMSE results achieved by SSA and SDD using noisy simulated data.

Simulation Run	RMSE		RRMSE ($\frac{SSA}{SDD}$)
	SSA	SDD	
1	2.72	4.87	0.55
2	2.71	4.83	0.56
3	2.69	3.98	0.54
4	2.72	4.68	0.58
5	2.73	4.96	0.54

3.2. The Bicode Data

3.2.1. Data Description

The expression level of *Bcd* gene in wild-type *Drosophila melanogaster* embryos is measured by using fluorescently tagged antibodies. The data used in this study is available via [32]. A complete explanation on the method, data, biological characteristics and different cleavage cycles can be found in [2,33,34].

3.2.2. The Results

Since a natural hint for grouping is the matrix of the absolute values of the *w*-correlations, calculated *w*-correlations for different time classes are presented in Table 3. If the absolute value of the *w*-correlations is small, then *Bcd* signal and its corresponding noise are almost *w*-orthogonal, but if it is large, then they are far from being *w*-orthogonal, and are therefore not very well separable. As it can be seen from the results, the *Bcd* signal can be effectively separated from the noise component, confirming the technique proposed in this paper performs well in noise reduction and pattern recognition.

Table 3. *w*-correlations between signal and residuals for *Bcd* gene expression series.

Time Class	<i>w</i> -Correlation	Time Class	<i>w</i> -Correlation
10	0.008	14(3)	0.001
11	0.005	14(4)	0.002
12	0.007	14(5)	0.001
13	0.003	14(6)	0.001
14(1)	0.001	14(7)	0.001
14(2)	0.001	14(8)	0.001

Figures 3 and 4 show the extracted trend obtained by SSA-MV. The black and red colours indicate original data and SSA trend, respectively. As it can be seen from Figures 3 and 4, as the series length increases the fluctuation in data increases, confirming more noise in AP position. Here we can see that the trend component is precisely extracted for all time classes, indicating the robustness of SSA

to outliers. This can be considered as another advantage of SSA over classical methods, including the parametric approach currently used for *Bcd* analysis.

Figure 3. Temporal dynamics of *Bcd* expression, cleavage cycles 10 to 13. Red and green lines show the trend extracted by SSA_{MV} and SDD , respectively.

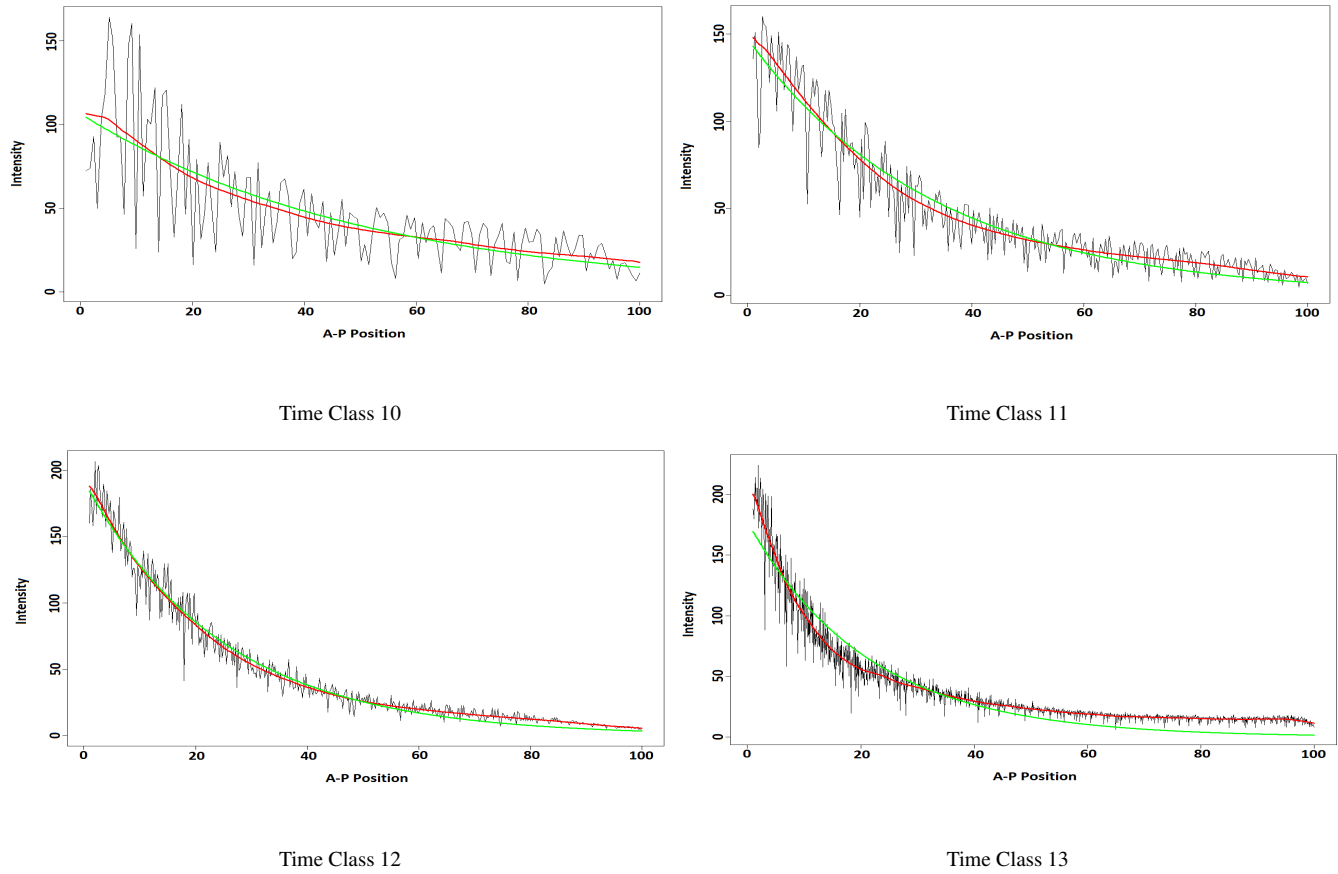
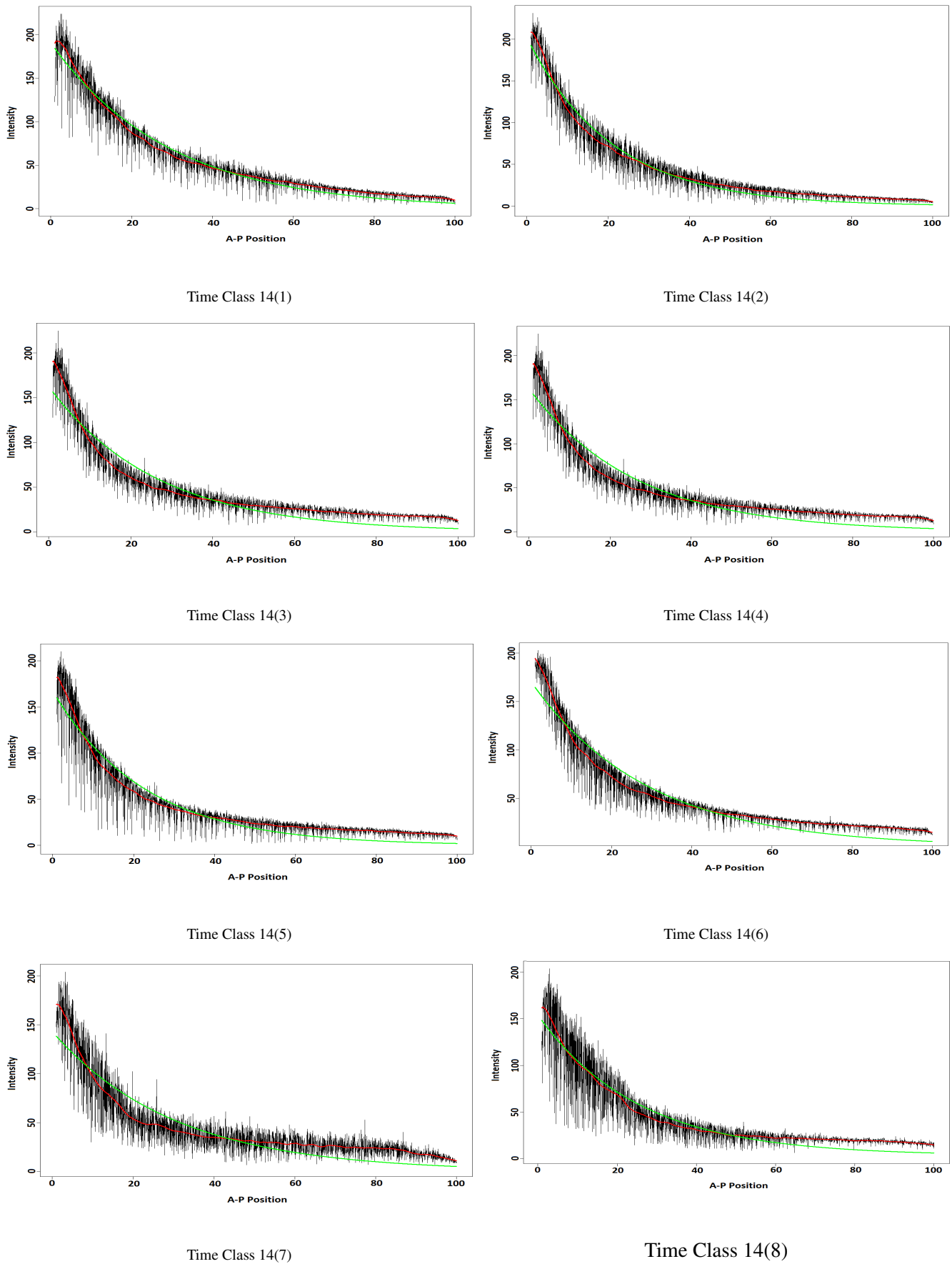


Figure 4. Temporal dynamics of *Bcd* expression, cleavage cycle 14. Red and green lines show the trend extracted by SSA_{MV} and SDD , respectively.



4. Conclusions

Extracting the *Bcd*'s signal was central to this study and the obtained results illustrate that the mentioned SSA filtering method used in this paper outperforms the previous methods used for filtering noisy *Bcd*. It was found that the SSA procedure is more flexible than the previous technique employed for the *Bcd* degradation modelling. The feasibility of capturing the trend of *Bcd* gene in *Drosophila* embryos suggests that our model may be of general use in evaluating other expressional systems. Thus, this method can be a valuable aid in analysing spatially inhomogeneous noisy data. In conclusion, the results confirm that filtering is very important for *Bcd* curve fitting and the SSA technique yields a promising result for *Bcd* analysis.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. How Instructors Request a Complimentary Sample. Available online: <http://highered.mcgraw-hill.com> (accessed on 30 June 2014).
2. Surkova, S.; Kosman, D.; Kozlov, K.; Manu; Myasnikova, E.; Samsonova, A.A.; Spirov, A.; Vanario-Alonso, C.E.; Samsonova, M.; Reinitz, J. Characterization of the *Drosophila* segment determination morphome. *Dev. Biol.* **2007**, *313*, 844–862.
3. Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **1969**, *27*, 1–47.
4. Nusslein-Volhard, C.; Wieschaus, E. Mutations affecting segment number and polarity in *Drosophila*. *Nature* **1980**, *287*, 795–801.
5. Liu, W.; Niranjana, M. Gaussian process modelling for bicoid mRNA regulation in spatio-temporal Bicoid profile. *Bioinformatics* **2011**, *28*, 366–372.
6. Holloway, D.M.; Harrison, L.G.; Kosman, D.; Vanario-Alonso, C.E.; Spirov, A.V. Analysis of pattern precision shows that *Drosophila* segmentation develops substantial independence from gradients of maternal gene products. *Dev. Dyn.* **2006**, *235*, 2949–2960.
7. Nusslein-Volhard, C.N. The bicoid morphogen papers (I): Account from CNV. *Cell* **2004**, *116*, 1–5.
8. Grimm, O.; Coppey, M.; Wieschaus, E. Modelling the bicoid gradient. *Development* **2010**, *137*, 2253–2264.
9. Liu, J.; He, F.; Ma, J. Morphogen gradient formation and action. *Fly* **2011**, *5*, 242–246.
10. Houchmandzadeh, B.; Wieschaus, E.; Leibler, S. Establishment of developmental precision and proportions in the early *Drosophila* embryo. *Nature* **2002**, *415*, 798–802.
11. Gregor, T.; Tank, D.W.; Wieschaus, E.F.; Bialek, W. Stability and Nuclear Dynamics of the Bicoid Morphogen Gradient. *Cell* **2007** *130*, 153–164.

12. Hilfinger, H.; Paulsson, J. Separating intrinsic from extrinsic fluctuations in dynamic biological systems. *Proc. Natl. Acad. Sci. USA* **2011**, *108*, 12167–12172.
13. Hassani, H.; Zokaei, M.; von Rosenc, D.; Amiri, S.; Ghodsi, M. Does noise reduction matter for curve fitting in growth curve models? *Comput. Methods Programs Biomed.* **2009**, *96*, 173–181.
14. Hassani, H.; Dionisio, A.; Ghodsi, M. The effect of noise reduction in measuring the linear and nonlinear dependency of financial markets. *Nonlin. Anal.: Real World Appl.* **2010**, *11*, 492–502.
15. Du, L.P.; Wu, S.H.; Liew, A.W.C.; Smith, D.K.; Yan, H. Spectral Analysis of Microarray Gene Expression Time Series Data of Plasmodium Falciparum. *Int. J. Bioinf. Res. Appl.* **2008**, *4*, 337–349.
16. Hassani, H. Singular spectrum analysis: Methodology and comparison. *J. Data Sci.* **2007**, *5*, 239–257.
17. Hassani, H.; Thomakos, D. A review on singular spectrum analysis for economic and financial time series. *Stat. Interface* **2010**, *3*, 377–397.
18. Hassani, H.; Soofi, S.; Zhigljavsky, A. Predicting Inflation Dynamics with Singular Spectrum Analysis. *J. R. Stat. Soc. Ser. A (Stat. Soc.)* **2012**, *176*, 743–760
19. Golyandina, N.E.; Holloway, D.M.; Lopesc, F.J.P.; Spirov, A.V.; Spirova, E.N.; Usevich, K.D. Measuring gene expression noise in early *Drosophila* embryos: nucleus-to-nucleus variability. *Int. Conf. Comput. Sci.* **2012**, *9*, 373–382.
20. Spirov, A.V.; Golyandina, N.E.; Holloway, D.M.; Alexandrov, T.; Spirova, E.N.; Lopes, F.J.P. Measuring Gene Expression Noise in Early *Drosophila* Embryos: The Highly Dynamic Compartmentalized Micro-environment of the Blastoderm Is One of the Main Sources of Noise. *Evol. Comput. Mach. Learn. Data Mining Bioinf.* **2012**, *7246*, 177–188.
21. Holloway, D.M.; Lopes, F.J.P.; da Fontoura Costa, L.; Travencolo, B.A.N.; Golyandina, N.; Usevich, K.; Spirov, A.V. Gene Expression Noise in Spatial Patterning: Hunchback Promoter Structure Affects Noise Amplitude and Distribution in *Drosophila* Segmentation. *PLoS Comput. Biol.* **2011**, *7*, e1001069.
22. Hassani, H. Singular spectrum analysis based on the minimum variance estimator. *Nonlin. Anal.: Real World Appl.* **2010**, *11*, 2065–2077.
23. Alexandrov, T.; Golyandina, N.; Sprov, A. Singular spectrum analysis of gene expression profiles of early drosophila embryo: Exponential-in-distance patterns. *Res. Lett. Signal Process* **2010**, *2008*, 5.
24. Ghodsi, M.; Hassani, H.; Sanei, S.; Hicks, Y. The use of noise information for detection of temporomandibular disorder. *Biomed. Signal Process. Control* **2010**, *4*, 79–85.
25. Ghodsi, M.; Hassani, H.; Sanei, S. Extracting fetal heart signal from noisy maternal ECG by multivariate singular spectrum analysis. *Stat. Interface* **2010**, *3*, 399–411.
26. Vivian, T.; Tang, Y.; Yan, H. Noise Reduction in Microarray Gene Expression Data Based on Spectral Analysis. *Int. J. Mach. Learning Cybern.* **2012**, *3*, 51–57.
27. Mamou, J.; Feleppa, E.J. Singular Spectrum Analysis Applied to Ultrasonic Detection and Imaging of Brachytherapy Seeds. *J. Acoust. Soc. Am.* **2007**, *121*, 1790–1801.

28. De Moor, B. The singular value decomposition and long and short spaces on noisy matrices. *IEEE Trans. Signal Process.* **1993**, *41*, 2826–2838.
29. Golyandina, N.; Nekrutkin, V.; Zhigljavsky, A. *Analysis of Time Series Structure: SSA and related techniques*. Chapman & Hall/CRC: Boca Raton, FL, USA, 2001.
30. Hassani, H.; Heravi, H.; Zhigljavsky, A. Forecasting European Industrial Production with Singular Spectrum Analysis, *Int. J. Forec.* **2009**, *25*, 103–118.
31. Van Huffel, S. Enhanced resolution based on minimum variance estimation and exponential data modeling, *Signal Process.* **1993**, *33*, 333–355.
32. A Database of Segmentation gene Expression in *Drosophila*. Available online: <http://urchin.spbcas.ru/flyex/> (accessed on 30 June 2014).
33. Pisarev, A.; Poustelnikova, E.; Samsonova, M.; Reinitz, J. FlyEx, the quantitative atlas on segmentation gene expression at cellular resolution. *Nucl. Acids Res.* **2008**, doi:10.1093/nar/gkn717.
34. Poustelnikova, E.; Pisarev, A.; Blagov, M.; Samsonova, M.; Reinitz J. A database for management of gene expression data *in situ*. *Bioinformatics* **2008**, *20*, 2212–2221.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).