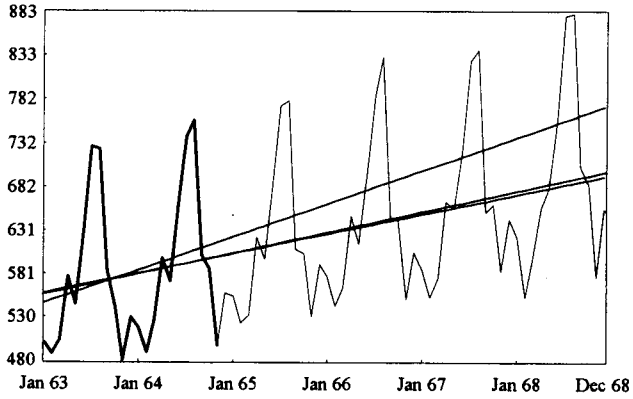

Figure 1.47 *Hotels: continuation of linear functions.*

Comparison of these lines shows that linear regression fails to find a good approximation of the main tendency of the whole series based on the first 23 points. On the other hand, the Double centring line is very close to the global linear regression line, but it uses only the first 23 points of the series rather than the full number 168.

Note that using Single or Double centring SSA, one can extract not only constants or linear components of time series. Other components of interest (such as oscillatory ones) can be extracted in the same manner as in Basic SSA. For example, for the series, containing the first 23 points of the 'Hotels' data and Double centring with $L = 12$ (see Fig. 1.46), the eigentriples 1-2, 3-4, 5-6, 7, 8-9 and 10-11 describe harmonics with $\omega = 1/12, 2/12, 3/12, 6/12, 4/12$ and $5/12$, respectively.

Moreover, if the time series has a general linear-like tendency, then the Double centring approach is often preferable to Basic SSA.

1.7.2 Stationary series and Toeplitz SSA

If the length N of the series F is not sufficiently large and the series is assumed to be stationary, then the usual recommendation is to replace the matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ by some other matrix, which takes into account the stationarity of the series.

Note first that we can consider the *lag-covariance matrix* $\mathbf{C} = \mathbf{S}/K$ instead of $\mathbf{S}$ for obtaining the SVD of the trajectory matrix $\mathbf{X}$. Indeed, the difference between the SVDs of the matrices $\mathbf{S}$ and $\mathbf{C}$ lies only in the magnitude of the corresponding eigenvalues (for $\mathbf{S}$ they are K times larger); the singular vectors of both matrices are the same. Therefore, we can use both $\mathbf{S}$ and $\mathbf{C}$ in Basic SSA with the same effect.

Denote by $c_{ij} = c_{ij}(N)$ the elements of the lag-covariance matrix $\mathbf{C}$. If the time series is stationary, and $K \to \infty$, then $\lim c_{ij} = R_f(|i-j|)$ as $N \to \infty$, where $R_f(k)$ stands for the lag k term of the time series covariance function; see Sections 1.4.1 and 6.4. (Recall that according to our agreement of Section 1.4.1, any infinite stationary series has zero average.)

Therefore, the main idea is to take the Toeplitz version of the lag-covariance matrix, that is to put equal values $\widetilde{c}_{ij}$ in each matrix diagonal $|i-j| = k$. Of course, the convergence $\widetilde{c}_{ij} \to R_f(|i-j|)$ must be kept.

There are several ways of getting the Toeplitz lag-covariance matrices from the series (see Elsner and Tsonis, 1996, Chapter 5.3). The main one is to use the standard estimate of the covariance function of the series and to transform it into an $L \times L$ matrix. More precisely (see Anderson, 1994, Chapter 8.2), for the time series $F = (f_0, \ldots, f_{N-1})$ and a fixed window length L, we take the matrix $\widetilde{\mathbf{C}}$ with the elements

$$\widetilde{c}_{ij} = \frac{1}{N-|i-j|} \sum_{m=0}^{N-|i-j|-1} f_m f_{m+|i-j|}, \quad 1 \le i,j \le L, \tag{1.32}$$

rather than Basic SSA lag-covariance matrix $\mathbf{C} = \mathbf{S}/K$ with the elements

$$c_{ij} = \frac{1}{K} \sum_{m=0}^{K-1} f_{m+i-1} f_{m+j-1}, \quad 1 \le i,j \le L. \tag{1.33}$$

Having obtained the *Toeplitz lag-covariance matrix* $\widetilde{\mathbf{C}}$ we calculate its orthonormal eigenvectors $H_1, \ldots, H_L$ and decompose the trajectory matrix:

$$\mathbf{X} = \sum_{i=1}^{L} H_i Z_i^{\mathrm{T}}, \tag{1.34}$$

where $Z_i = \mathbf{X}^{\mathrm{T}} H_i$. We thus obtain an orthogonal matrix decomposition of the kind discussed in Section 4.2.1. Setting $\lambda_i = ||Z_i||^2$ and $Q_i = Z_i/\sqrt{\lambda_i}$ (here we formally assume that $\widetilde{\mathbf{C}}$ has full rank), we come to the decomposition of the trajectory matrix $\mathbf{X}$ into a sum similar to the usual SVD. The grouping and diagonal averaging can then be made in the standard way. Note that the numbers λ_i (which may be called squared *Toeplitz singular values*) generally do not coincide with the eigenvalues of the matrix $\widetilde{\mathbf{C}}$.

If the initial series is a sum of a constant series with the general term c_0 and a stationary series, then centring seems to be a convenient procedure (since we are dealing with finite time series, the centring can be applied for $c_0 = 0$ as well). One way is to centre the entire series before calculating the matrix (1.32).

The other method is to apply the Single centring. For *Toeplitz SSA* with the lag-covariance matrix (1.32) this means that we extract the product

$$M_{ij} = \left(\frac{1}{n(i,j)} \sum_{m=0}^{n(i,j)-1} f_m \right) \left(\frac{1}{n(i,j)} \sum_{m=0}^{n(i,j)-1} f_{m+|i-j|} \right)$$

(here we used the notation $n(i,j) = N - |i-j|$) from $\widetilde{c}_{ij}$, find the eigenvectors $H_1, \ldots, H_L$ of the above matrix, compute the (single) centred trajectory matrix $\mathbf{X}^*$ as was described in Section 1.7.1, obtain $Z_i = (\mathbf{X}^*)^{\mathrm{T}} H_i$, and come to the decomposition similar to (1.34) with an additional matrix term $\mathbf{A}$ corresponding to the Single centring. Note that unlike Basic SSA, the Toeplitz SSA is not invariant with respect to the substitution of $K = N - L + 1$ for the window length L, even without centring.

The Toeplitz construction of the lag-covariance matrix seems to have an advantage since the matrix elements (1.32) are generally closer than (1.33) to the terms $R_f(|i-j|)$ of the theoretical covariance function, due to a wider range of averaging. Nevertheless, it is not universally better since we are not dealing with the lag-covariance matrix itself but rather with some specific features of the decompositions of the trajectory matrices, such as separability.

First, the Toeplitz SSA is not aimed at nonstationary series. If the series has a strong nonstationary component, then Basic SSA seems to be preferable. For example, if we are dealing with a pure exponential series, then it is described by a single eigentriple (see Sections 1.6.1 and 5.1 for details) for any window length, while Toeplitz SSA produces L eigentriples for window length L with harmonic-like eigenvectors. The same effect takes place for the linear series, exponential-cosine series, etc. In terms of Section 4.2.1, Toeplitz SSA often produces a decomposition, which is not minimal.

Second, Toeplitz SSA generally produces a nonoptimal decomposition. The decomposition of the trajectory matrix produced by SVD (it is used in Basic SSA and Single and Double centring SSA) is optimal in the the sense that each eigenvalue is the solution of a certain optimization problem; in other words, each eigenvalue is as large as it can be. Therefore, the main series effects are described by the leading SVD eigentriples, but even subsequent eigentriples can be meaningful.

If we have nonoptimal orthogonal decomposition of the trajectory matrix, it is more 'spread' and the problem similar to the problem of small 'almost equal singular values' becomes even more serious.

Moreover, for long stationary series, both methods give practically the same results. Yet, for relatively short stationary and noisy series, Toeplitz SSA can be advantageous.

Example 1.7 *'Tree rings': four modulated harmonics*
Let us consider the 'Tree rings' example (see Section 1.3.2). The periodogram (Fig. 1.48) of the series shows four sharp peaks corresponding approximately to the periods $T_1 = 74, T_2 = 52, T_3 = 42$ and $T_4 = 12.5$.

If we take the window length $L = 334$, then Basic and Single centring SSA (both using SVD) extract periodicities corresponding to T_1 and T_4 but produce a mixture of the two other periodic components. Standard Toeplitz SSA with Single centring works better (see Fig. 1.49) and extracts all the leading periodicities at once.

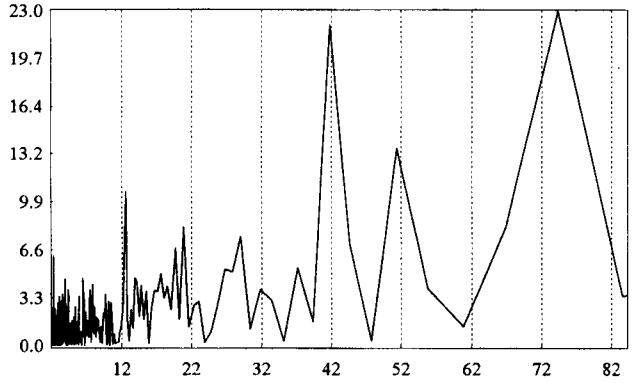

Figure 1.48 *Tree rings: periodogram in periods up to* $T = 85$.

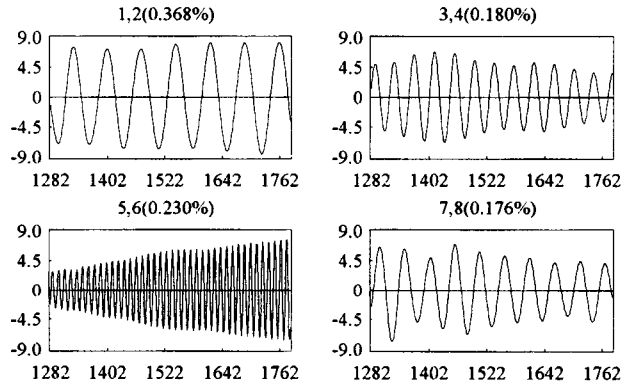

Figure 1.49 *Tree rings: four periodic components. First 500 points.*

Remark 1.1 In Example 1.7 we did not discuss whether the extracted periodicities are the true ones or produced by the aperiodic component of the series (see Section 6.4). Our aim was to demonstrate their extraction.

1.7.3 Close singular values

As was discussed in Section 1.6.2, close singular values of SVD cause difficulties that are difficult to overcome by modifying the window length L. Nevertheless, there are several techniques that can help to solve the problem. Let us discuss two of them.

(a) Series modification

Sometimes one can modify the series in such a manner that this problem disappears. The theoretical base of such effects is the following simple fact: if the time series $F^{(1)}$ and $F^{(2)}$ are weakly separable, then for a wide range of constants $c \neq 0$ the time series $F^{(1)}$ and $cF^{(2)}$ are strongly separable. In practice, we use this fact in some approximate sense.

Let us consider two examples of this kind.

1. If we want to extract a small slowly varying trend whose components are mixed with other series components, then it can be worthwhile to add a constant to the series and use a relatively small window length for the trend extraction. Then the new trend will be described by the leading eigentriple, and there will be no problem in its extraction. The added constant has to be subtracted from the extracted series.

   The example 'England temperatures' (Section 1.6.2) is of this kind if we deal with it in the reverse manner; being centred, the time series is complex for the rough trend extraction, but if we add to the centred series a constant, equal to 9.18 (that is, if we come back to the uncentered data), a rather wide range of window lengths will provide the extraction.

2. Assume that our aim is extracting a harmonic with a known frequency ω and this harmonic is mixed with some other time series components due to their close singular values. If the selected window length L provides a weak separability of the harmonic of interest, then we can add a harmonic of the same frequency (and some amplitude and phase) to the series. Under the proper choice of these parameters, the singular values corresponding to the harmonic will be enlarged enough so that they will not be mixed with any other series components (for example, the harmonic will be described by the leading eigentriples). Therefore, the modified harmonic will be easily extracted.

Example 1.8 *'Rosé wine': adding a harmonic component*
To illustrate the extraction of a harmonic component from the series, let us consider the example 'Rosé wine' described in Section 1.4.1 (see Fig. 1.17 for the time series and Fig. 1.18 for its periodogram). As was mentioned in Section 1.6, the harmonics with frequencies 4/12 and 5/12 are mixed under the choice $L = 84$. Moreover, other window lengths lead to mixing of other harmonics due to a complex nonstationary structure of the series.

However, if we add to the series a harmonic with frequency 4/12 (that is, period 3), zero phase and amplitude 30, then the new quarterly harmonics will be extracted under the choice of the same $L = 84$ and the pair of the second and third eigentriples. The final result is obtained by subtracting the additional harmonic component.

Note that the problem of close singular values can be solved by other modifications of Basic SSA as well. For instance, the Toeplitz SSA helps in extracting harmonic components in the 'Tree rings' example of Section 1.7.2. However, this

example seems to be a good illustration of the advantages of a concrete technique related to the problem of close singular values rather than an illustration of the absolute advantage of the Toeplitz SSA.

(b) Sequential SSA

The mixing problem of the time series components (formally, the problem of close singular values for weakly separable series components) may be resolved in one more manner, by the so-called *Sequential SSA*.

The two-step Sequential SSA can be described as follows. First, we extract several time series components by Basic SSA with a certain window length L_1. Then we apply Basic SSA to the residuals and extract several series components once again. The window length L_2 of the second stage is generally different from L_1.

Having extracted two sets of time series components, we can group them in different ways. For instance, if a rough trend has been extracted at the first stage and other trend components at the second stage, then we have to add them together to obtain the accurate trend.

Let us illustrate this by an example.

Example 1.9 *Long 'Unemployment' series: extraction of harmonics*
Consider the 'Unemployment' series starting from January 1948 (note that 'Unemployment' example of Section 1.3.6 has April 1950 as its starting point). The series is depicted in Fig. 1.50 (thin line).

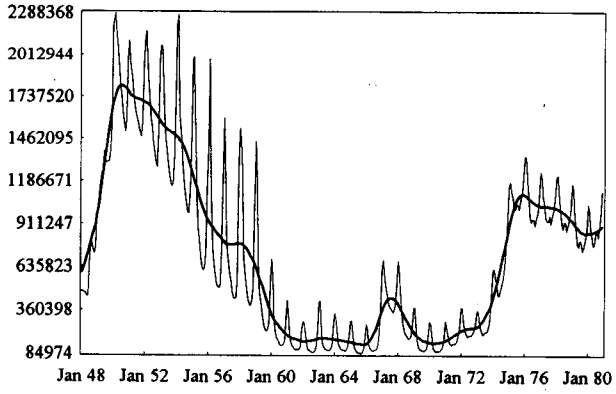

Figure 1.50 *Long 'Unemployment' series: time series from January 1948.*

Comparing Fig. 1.50 with Fig. 1.12, we see that the trend of the long 'Unemployment' series (thick line) has a more complex structure than that of the shorter one. Selection of a large window length would mix the trend and periodic components of the series. For small window lengths the periodic components are not

separable from each other, and therefore these lengths are not suitable. Hence, Basic SSA fails to extract (amplitude-modulated) harmonic components of the series.

The two-stage Sequential SSA proves to be a better method in this case. If we apply Basic SSA with $L = 12$ to the initial series, then the first eigentriple will describe the trend, which is extracted rather well: the trend component does not include high frequencies, while the residual component practically does not contain low ones (see Fig. 1.51 for the residual series).

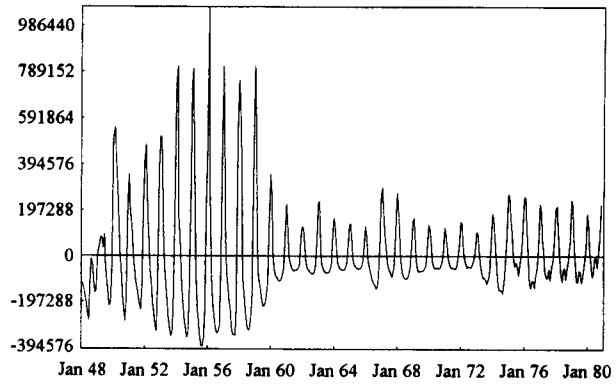

Figure 1.51 *Long 'Unemployment' series: trend residuals.*

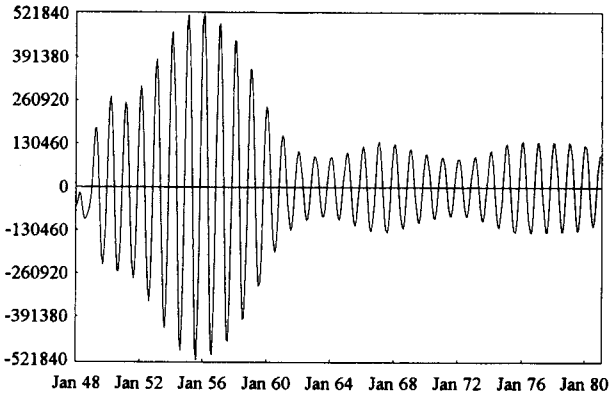

Figure 1.52 *Long 'Unemployment' series: annual periodicity.*

The second Sequential SSA stage is applied to the residual series with $L = 180$. Since the series is amplitude modulated, the main periodogram frequencies

(annual $\omega = 1/12$, half-annual $\omega = 1/6$ and 4-months $\omega = 1/4$) are somewhat spread out, and therefore each (amplitude-modulated) harmonic can be described by several (more than 2) eigentriples.

Periodogram analysis of the obtained singular vectors shows that the leading 14 eigentriples with share 91.4% can be related to 3 periodicities: the eigentriples $1, 2, 5-8, 13, 14$ describe the annual amplitude-modulated harmonic (Fig. 1.52), the eigentriples $3, 4, 11-12$ are related to half-year periodicity, and the eigentriples $9, 10$ describe the 4-months harmonic.

The same technique can be applied to the 'Births' series if we want to obtain better results than those described in Section 1.3.4. (See Section 1.6.2 for a discussion concerning the large window length problem in this example.)

1.7.4 Envelopes of highly oscillating signals

The capabilities of SSA in separating signals with high and low frequencies can be used in a specific problem of enveloping highly oscillating sequences with slowly varying amplitudes. The simple idea of such a technique can be expressed as follows.

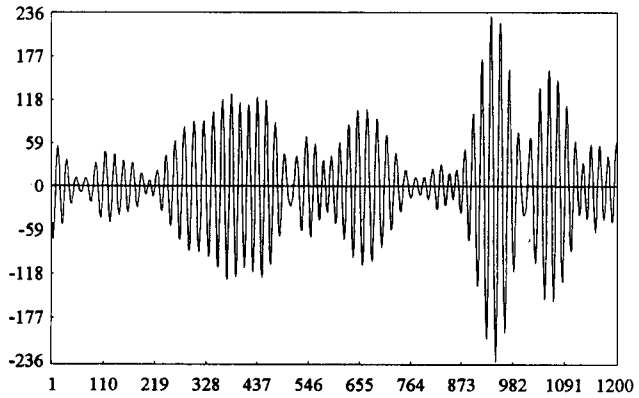

Figure 1.53 *EEG: α-rhythm. First 1200 points.*

Let $f_n = A(n)\cos(2\pi\omega n)$ where ω is large and $A(n)$ is slowly varying. Then

$$g_n \stackrel{\text{def}}{=} 2f_n^2 = A^2(n) + A^2(n)\cos(4\pi\omega n). \tag{1.35}$$

Since $A^2(n)$ is slowly varying and the second term on the right-hand side of (1.35) oscillates rapidly, one can gather the slowly varying terms of the SSA decomposition for g_n, and therefore approximately extract the term $A^2(n)$ from the series (1.35). All we need to do then is to take the square root of the extracted term.

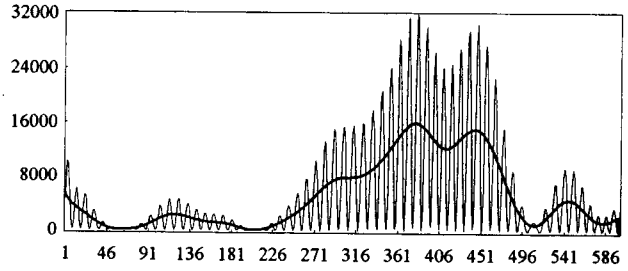

Figure 1.54 *EEG: series G and its slowly varying component. First 600 points.*

Example 1.10 *EEG: envelope of α-rhythm*
This idea is illustrated by the time series F representing an α-rhythm component of an electroencephalogram (EEG). The whole series F consists of approximately 3500 points; its first 1200 points can be seen in Fig. 1.53. The series can be described as an amplitude-modulated harmonic with the main frequency approximately equal to $1/20$.

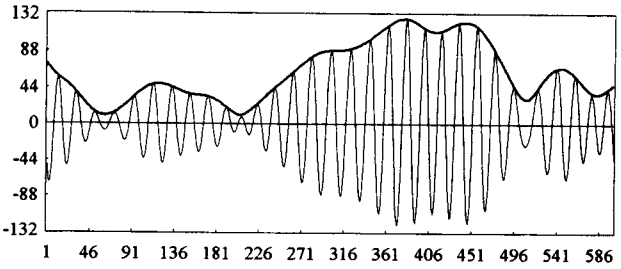

Figure 1.55 *EEG: α-rhythm and its envelope. First 600 points.*

Let us consider the square of the initial series multiplied by 2 and denote it by G. Taking window length $L = 60$ and reconstructing the low-frequency part of the time series G from the eigentriples $1, 4, 7$ and 10, we obtain an estimate of $A^2(n)$ (the first 600 points of the reconstructed series are depicted in Fig. 1.54 by the thick line; the thin line corresponds to the series G).

By taking the square root of the estimate we obtain the result. (See Fig. 1.55, where the first 600 points of the initial series with its envelope are depicted.)

It may be interesting to note that the α-rhythm time series under consideration was extracted from the initial EEG signal by 5-stage Sequential SSA with different window lengths (the largest was equal to 600).

Note also that to obtain the resulting envelope we may need some smoothing to remove very small but existing parts of highly oscillating components. As usual, Basic SSA with small window length would do the job.

CHAPTER 2

SSA forecasting

A reasonable forecast of a time series can be performed only if the following conditions are met:

1. The series has a structure.
2. A mechanism (method, algorithm) identifying this structure is found.
3. A method of the time series continuation, based on the identified structure, is available.
4. The structure of the time series is preserved for the future time period over which we are going to forecast (continue) the series.

All these conditions are natural. Of course, condition 4 cannot be validated with the help of the data to be forecasted. Moreover, the structure of the series can hardly be identified uniquely (for example, if the series has a noise component). Therefore, the situation of different (and even 'contradictory') forecasts is not impossible. Thus, it is important not only to realize and indicate the structure under continuation, but also to check its stability.

At any rate, a forecast can be made only if a certain model is built. The model can either be derived from the data or at least checked against the data. In SSA forecasting, these models can be described with the help of the linear recurrent formulae (or equations). Note that in general the dimension (in other words, the order) of the recurrent formulae may be unknown.

The class of series governed by linear recurrent formulae (LRFs) is rather wide and important for practical implications. For instance, an infinite series is governed by some LRF if and only if it can be represented as a linear combination of products of exponential, polynomial and harmonic series. (See Chapter 5 for a review of the entire theory.)

The series governed by LRFs admits natural *recurrent continuation* since each term of such a series is equal to a linear combination of several preceding terms. Of course, the coefficients of this linear combination can be used for the continuation as well.

It is important that we need not necessarily search for an LRF of minimal dimension. Indeed, any other LRF governing the series produces the same continuation.

The theory of Section 5.2, Chapter 5, indicates how to find an LRF, which governs a series, with the help of SSA. The general idea can be described as follows.

Let d be the minimal dimension of all LRFs governing F. (In this case we shall say that the time series F is governed by a minimal LRF of dimension d.) It can be proved that if the window length L is larger than d, and the length of the series is sufficiently large, then the trajectory space of the series F is d-dimensional. Basic SSA provides a natural basis for the trajectory space.

The trajectory space determines (under mild and natural restrictions) an LRF of dimension $L-1$ that governs the series. If we apply this LRF to the last terms of the initial series F, we obtain the continuation of F.

The same idea may work if we want to continue an additive component $F^{(1)}$ of a series F. Here we assume that $F^{(1)}$ is governed by an LRF and is strongly separable from the residual series $F^{(2)} = F - F^{(1)}$ for the selected value of the window length L. It should be mentioned that if two series are strongly separable, then each of them must satisfy some LRF (see Remark 6.1 in Section 6.1.1).

In practice, it is not reasonable to assume that the series of interest is governed by an LRF of relatively small dimension. In this way we come to the concept of *approximate recurrent continuation*, which can and will also be called the *recurrent forecasting*. We thus suppose that the series F under consideration can be expressed as a sum of the series $F^{(1)}$ admitting recurrent continuation and the residual series $F^{(2)}$. If we consider the residuals as a noise, then we have the problem of forecasting the signal $F^{(1)}$ in the presence of the noise $F^{(2)}$. We may also have the problems of forecasting the series $F^{(1)}$ regarded as a trend or a seasonal component of F.

The main assumption is that for a certain window length L, the series components $F^{(1)}$ and $F^{(2)}$ are approximately strongly separable. Then, acting as in Basic SSA, we reconstruct the series $F^{(1)}$ with the help of a selected set of eigentriples and obtain approximations to both the series $F^{(1)}$ and its trajectory space. In other words, we obtain both the LRF, approximately governing $F^{(1)}$, and the initial data for this formula. Hence we obtain a forecast of the series $F^{(1)}$.

The theory of the method can be found in Chapter 5. The contents of the present Chapter are as follow.

Section 2.1 formally describes the general SSA forecasting algorithm. The rest of the chapter is devoted to study of this algorithm and related discussions.

Section 2.2 describes the principles of SSA forecasting and its relations to linear recurrent formulae. Several modifications of the general SSA forecasting algorithm are considered in Section 2.3.

Section 2.4 is devoted to a description of different ways of constructing confidence intervals that can be used for checking the forecast accuracy and stability. After the summarizing Section 2.5, several forecasting examples are presented in Section 2.6.

When dealing with continuation, we always need to bear in mind the length of the series under continuation. Therefore, we usually incorporate this length into the notation of the series and write, for example, F_N rather than simply F.

2.1 SSA recurrent forecasting algorithm

Let us formally describe the forecasting algorithm under consideration.

Algorithm inputs:

(a) Time series $F_N = (f_0, \ldots, f_{N-1})$, $N > 2$.

(b) Window length L, $1 < L < N$.

(c) Linear space $\mathfrak{L}_r \subset \mathbf{R}^L$ of dimension $r < L$. It is assumed that $e_L \notin \mathfrak{L}_r$, where $e_L = (0, 0, \ldots, 0, 1)^{\mathrm{T}} \in \mathbf{R}^L$. In other terms, $\mathfrak{L}_r$ is not a 'vertical' space. In practice, the space $\mathfrak{L}_r$ is defined by its certain orthonormal basis, but the forecasting results do not depend on this concrete basis.

(d) Number M of points to forecast for.

Notations and Comments:

(a) $\mathbf{X} = [X_1 : \ldots : X_K]$ (where $K = N - L + 1$) is the trajectory matrix of the time series F_N.

(b) $P_1, \ldots, P_r$ is an orthonormal basis in $\mathfrak{L}_r$.

(c) $\widehat{\mathbf{X}} \stackrel{\text{def}}{=} [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} P_i P_i^{\mathrm{T}} \mathbf{X}$. The vector $\widehat{X}_i$ is the orthogonal projection of X_i onto the space $\mathfrak{L}_r$.

(d) $\widetilde{\mathbf{X}} = \mathcal{H}\widehat{\mathbf{X}} = [\widetilde{X}_1 : \ldots : \widetilde{X}_K]$ is the result of the Hankelization of the matrix $\widehat{\mathbf{X}}$. The matrix $\widetilde{\mathbf{X}}$ is the trajectory matrix of some time series $\widetilde{F}_N = (\widetilde{f}_0, \ldots, \widetilde{f}_{N-1})$.

(e) For any vector $Y \in \mathbf{R}^L$ we denote by $Y_\triangle \in \mathbf{R}^{L-1}$ the vector consisting of the last $L-1$ components of the vector Y, while $Y^\nabla \in \mathbf{R}^{L-1}$ is the vector consisting of the first $L-1$ components of Y.

(f) We set $\nu^2 = \pi_1^2 + \ldots + \pi_r^2$, where π_i is the last component of the vector P_i ($i = 1, \ldots, L$). Since ν^2 is the squared cosine of the angle between the vector e_L and the linear space $\mathfrak{L}_r$, it can be called the *verticality coefficient* of $\mathfrak{L}_r$.

(g) Suppose that $e_L \notin \mathfrak{L}_r$. (In other words, we assume that $\mathfrak{L}_r$ is not a vertical space.) Then $\nu^2 < 1$. It can be proved (see Chapter 5, Theorem 5.2) that the last component y_L of any vector $Y = (y_1, \ldots, y_L)^{\mathrm{T}} \in \mathfrak{L}_r$ is a linear combination of the first components $y_1, \ldots, y_{L-1}$:

$$y_L = a_1 y_{L-1} + a_2 y_{L-2} + \ldots + a_{L-1} y_1.$$

Vector $\mathcal{R} = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ can be expressed as

$$\mathcal{R} = \frac{1}{1-\nu^2} \sum_{i=1}^{r} \pi_i P_i^\nabla \tag{2.1}$$

and does not depend on the choice of a basis $P_1, \ldots, P_r$ in the linear space $\mathfrak{L}_r$.

SSA recurrent forecasting algorithm:

In the above notations, define the time series $G_{N+M} = (g_0, \ldots, g_{N+M-1})$ by

the formula

$$g_i = \begin{cases} \widetilde{f}_i & \text{for } i = 0, \ldots, N-1, \\ \sum\limits_{j=1}^{L-1} a_j g_{i-j} & \text{for } i = N, \ldots, N+M-1. \end{cases} \tag{2.2}$$

The numbers $g_N, \ldots, g_{N+M-1}$ form the M terms of the SSA recurrent forecast. For brevity, we call this algorithm *SSA R-forecasting algorithm.*

Remark 2.1 Let us define the linear operator $\mathcal{P}^{(r)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ by the formula

$$\mathcal{P}^{(r)} Y = \begin{pmatrix} Y_\Delta \\ \mathcal{R}^{\mathrm{T}} Y_\Delta \end{pmatrix}, \quad Y \in \mathfrak{L}_r. \tag{2.3}$$

If setting

$$Z_i = \begin{cases} \widetilde{X}_i & \text{for } i = 1, \ldots, K, \\ \mathcal{P}^{(r)} Z_{i-1} & \text{for } i = K+1, \ldots, K+M, \end{cases} \tag{2.4}$$

the matrix $\mathbf{Z} = [Z_1 : \ldots : Z_{K+M}]$ is the trajectory matrix of the series G_{N+M}. Therefore, (2.4) can be regarded as the vector form of (2.2).

If $\mathfrak{L}_r$ is spanned by certain eigenvectors corresponding to the SVD of the trajectory matrix of the series F_N, then the corresponding SSA R-forecasting algorithm will be called the *Basic SSA R-forecasting algorithm.*

Remark 2.2 Denote by $\mathfrak{L}^{(L)} = \mathrm{span}(X_1, \ldots, X_K)$ the trajectory space of the series F_N. Suppose that $\dim \mathfrak{L}^{(L)} = r < L$ and $e_L \notin \mathfrak{L}^{(L)}$. If we use the Basic SSA R-forecasting algorithm with $\mathfrak{L}_r = \mathfrak{L}^{(L)}$, then $\mathbf{X} = \widehat{\mathbf{X}} = \widetilde{\mathbf{X}}$ and therefore $\widetilde{F}_N = F_N$. This means that the initial points $g_{N-L+1}, \ldots, g_{N-1}$ of the forecasting recurrent formula (2.2) coincide with the last $L-1$ terms of the series F_N.

2.2 Continuation and approximate continuation

The algorithmic scheme described in the previous section is related to both the series, which are governed by the linear recurrent formulae, and the SSA methodology. Let us describe the ideas that lead to SSA forecasting.

2.2.1 Linear recurrent formulae and their characteristic polynomials

The theory of the linear recurrent formulae and associated characteristic polynomials is well known (for example, Gelfond, 1967, Chapter V, §4). However, we provide here a short survey of the most essential results. A more formal description can be found in Chapter 5.

(a) Series governed by linear recurrent formulae

By definition, a nonzero series $F_N = (f_0, \ldots, f_{N-1})$ is governed by a *linear recurrent formula* (*LRF*) of dimension not exceeding $d \geq 1$ if

$$f_{i+d} = \sum_{k=1}^{d} a_i f_{i+d-k} \tag{2.5}$$

for certain $a_1, \ldots, a_d$ with $a_d \neq 0$ and $0 \leq i \leq N - d + 1$. In the notation of Section 5.2 this is expressed as $\mathrm{fdim}(F_N) \leq d$. If

$$d = \min(k : \mathrm{fdim}(F_N) \leq k),$$

then we write $\mathrm{fdim}(F_N) = d$ and call d the *finite-difference dimension* of the series F_N. In the case when F_N is governed by LRF (2.5) and $d = \mathrm{fdim}(F_N)$, the formula (2.5) is called *minimal*.

If (2.5) holds but we do not require that $a_d \neq 0$, then the time series F_N *satisfies* the LRF (2.5).

The class of series governed by LRFs is rather wide: it contains harmonic, exponential and polynomial series and is closed under term-by-term addition and multiplication. For instance, the exponential series $f_n = e^{\alpha n}$ is governed by the LRF $f_n = a f_{n-1}$ with $a = e^{\alpha}$, the harmonic series $f_n = \cos(2\pi\omega n + \phi)$ satisfies the equation

$$f_n = 2\cos(2\pi\omega)\, f_{n-1} - f_{n-2},$$

and so on. Other examples, as well as theoretical results, can be found in Section 5.2.

The difference between minimal and arbitrary LRFs governing the same series can be illustrated by the following example. For the exponential series F_N with $f_n = a^n$, $a = e^{\alpha}$ and $N \geq 3$, the LRF $f_n = a f_{n-1}$ is the minimal one and $\mathrm{fdim}(F_N) = 1$. On the other hand, the series $f_n = a^n$ satisfies the equation $f_n = 2a f_{n-1} - a^2 f_{n-2}$ for $2 \leq n \leq N - 1$.

To understand whether the LRF (2.5) is minimal for the series F_N with sufficiently large N, one can apply the following procedure. Consider the window length L $(1 < L < N)$ and suppose that $d < \min(L, K)$. In view of (2.5), the L-lagged vectors $X_1, \ldots, X_K$ satisfy the vector recurrent equation

$$X_{i+d} = \sum_{k=1}^{d} a_i X_{i+d-k}, \quad 1 \leq i \leq K - d.$$

Therefore, each X_i is a linear combination of $X_1, \ldots, X_d$. If these vectors are linearly independent, then the LRF (2.5) is minimal and vice versa.

These assertions can be formulated in other terms. Denote by $\mathfrak{L}^{(L)}$ the trajectory space of the series F_N satisfying (2.5). If $d < \min(L, K)$, then the equalities $\mathrm{fdim}(F_N) = d$ and $\dim \mathfrak{L}^{(L)} = d$ are equivalent. Such a reformulation leads to a new concept.

Let $1 \le d \le L$. By definition, an arbitrary series F_N has *L-rank* d (i.e., $\mathrm{rank}_L(F_N) = d$) if $\dim \mathfrak{L}^{(L)} = d$.

If $\mathrm{rank}_L(F_N) = d$ for any L such that $d < \min(L, K)$, then the time series F_N has *rank* d (briefly, $\mathrm{rank}(F_N) = d$).

Roughly speaking, each time series F_N with $\mathrm{fdim}(F_N) = d$ has rank, and this rank is equal to d. The following simple example shows that the opposite assertion is not true: let us take $N = 7$ and $F_N = (1, 1, 1, 1, 1, 1, 2)$; then for each $L = 2, \ldots, 6$ we have $\mathrm{rank}_L(F_N) = 2$, while no LRF of dimension $d < 6$ can govern this series.

However, if $\mathrm{rank}_L(F_N) = d < L$, then the series F_N (with the exception of several first and last terms) is governed by an LRF of dimension $d_0 \le d$.

This LRF can be found by the procedure described in Theorem 5.1 of Chapter 5, but the procedure seems to be difficult for practical computations.

Moreover, let $L > \mathrm{rank}_L(F_N)$ and $e_L \notin \mathfrak{L}^{(L)}$. Let us denote $r = \dim \mathfrak{L}^{(L)}$ and take $\mathfrak{L}_r = \mathfrak{L}^{(L)}$. Then, as shown in Theorem 5.2 in the same chapter, the series F_N satisfies the LRF

$$f_{L+i-1} = a_1 f_{L+i-2} + \ldots + a_{L-1} f_i, \quad 0 \le i \le K-1, \tag{2.6}$$

where $\mathcal{R} = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ is defined in (2.1).

This fact has a purely geometric origin; due to Theorem 5.2, if $\mathfrak{L} \subset \mathbf{R}^L$ is a linear subspace of dimension $r < L$ and $e_L \notin \mathfrak{L}$, then the last component y_L of any vector $Y \in \mathfrak{L}$ is equal to the inner product $\mathcal{R}^{\mathrm{T}} Y^{\nabla}$, where the vector $Y^{\nabla} \in \mathbf{R}^{L-1}$ consists of the first $L-1$ components of the vector Y and $P_1, \ldots, P_r$ is an orthonormal basis of $\mathfrak{L}$.

(b) Characteristic polynomials and their roots

Let the series $F_N = (f_0, \ldots, f_{N-1})$ have finite-difference dimension d and is governed by the LRF

$$f_{d+i} = a_1 f_{d+i-1} + a_2 y_{d+i-2} + \ldots + a_d y_i, \qquad a_d \ne 0, \tag{2.7}$$

for $0 \le i \le N-d$. Consider the *characteristic polynomial* of the LRF (2.7):

$$P_d(\lambda) = \lambda^d - \sum_{k=1}^{d} a_k \lambda^{d-k}.$$

Let $\lambda_1, \ldots, \lambda_p$ be the different (complex) roots of the polynomial $P_d(\lambda)$. Since $a_d \ne 0$, these roots are not equal to zero. We also have $k_1 + \ldots + k_p = d$, where k_m are the multiplicities of the roots λ_m $(m = 1, \ldots, p)$.

Denote $f_n(m, j) = n^j \lambda_m^n$ for $1 \le m \le p$ and $0 \le j \le k_m - 1$. Theorem 5.3 of Section 5.2 tells us that the general solution of the equation (2.7) is

$$f_n = \sum_{m=1}^{p} \sum_{j=0}^{k_m - 1} c_{mj} f_n(m, j), \tag{2.8}$$

with certain complex c_{mj}. The specific values of the c_{mj} are defined by the first d elements of the series F_N: $f_0, \ldots, f_{d-1}$.

Thus, each root λ_m produces a component

$$f_n^{(m)} = \sum_{j=0}^{k_m-1} c_{mj} f_n(m, j) \tag{2.9}$$

of the series f_n.

Let us fix m and consider this component in the case $k_m = 1$, which is the main case in practice. Set $\lambda_m = \rho e^{i2\pi\omega}$, $\omega \in (-1/2, 1/2]$, where $\rho > 0$ is the modulus (absolute value) of the root and $2\pi\omega$ is its polar angle.

If ω is either 0 or $1/2$, then λ_m is a real root of the polynomial $P_d(\lambda)$ and the series component $f_n^{(m)}$ is real and is equal to $c_{m0}\lambda_m^n$. This means that $f_n^{(m)} = A\rho^n$ for positive λ_m and $f_n^{(m)} = A(-1)^n\rho^n = A\rho^n \cos(\pi n)$ for negative λ_m. The latter case corresponds to the exponentially modulated saw-tooth sequence.

All other values of ω lead to complex λ_m. In this case, P_d has a complex conjugate root $\lambda_l = \rho e^{-i2\pi\omega}$ of the same multiplicity $k_l = 1$. We thus can assume that $0 < \omega < 1/2$ and describe a pair of conjugate roots by the pair of real numbers (ρ, ω) with $\rho > 0$ and $\omega \in (0, 1/2)$.

If we add together the components $f_n^{(m)}$ and $f_n^{(l)}$ corresponding to these conjugate roots, then we obtain the real series $A\rho^n \cos(2\pi\omega n + \phi)$ with A and ϕ expressed in terms of c_{m0} and c_{l0}.

The asymptotic behaviour of $f_n^{(m)}$ essentially depends on $\rho = |\lambda_m|$. Let us consider the simplest case $k_m = 1$ as above. If $\rho < 1$, then $f_n^{(m)}$ rapidly tends to zero and asymptotically has no influence on the whole series (2.8). Alternatively, the root with $\rho > 1$ and $|c_{m0}| \neq 0$ leads to a rapid increase of $|f_n|$ (at least for a certain subsequence of n).

For example, if $\lambda_m = \rho = 0.8$ and $|c_{m0}| \neq 0$, then $|f_n^{(m)}|$ becomes smaller by approximately a factor 10 in 10 time steps and by a factor $5 \cdot 10^9$ in 100 steps. If $\lambda_m = \rho = 1.2$ (and $|c_{m0}| \neq 0$), then $|f_n^{(m)}|$ is increased approximately 6-fold in 10 time steps and $8 \cdot 10^7$-fold in 100 steps. Similar effects hold for the series component $A\rho^n \cos(2\pi\omega n + \phi)$ corresponding to a pair of conjugate complex roots: the series amplitude $A\rho^n$ rapidly decreases or increases depending on the inequalities $\rho < 1$ or $\rho > 1$.

The root λ_m with $k_m > 1$ produces k_m terms in the sum (2.9). For example, if $\lambda_m = 1$ and $k_m = 2$, then $f_n^{(m)} = An + B$ for some A and B. In other words, the root 1 of multiplicity 2 generates a linear series. Example 5.10 of Section 5.2 treats the general case $k_m = 2$ in detail.

If the series F_N has finite-difference dimension d, then the characteristic polynomial of its minimal LRF (2.7) has d roots. As was mentioned above, the same series satisfies many other LRFs of certain dimensions $r > d$. Consider such an

LRF

$$f_{r+i} = b_1 f_{r+i-1} + b_2 y_{r+i-2} + \ldots + b_r y_i. \tag{2.10}$$

The characteristic polynomial $P_r(\lambda)$ of the LRF (2.10) has r roots with d roots (we call them the *main roots*) coinciding with the roots of the minimal LRF. The other $r-d$ roots are *extraneous*: in view of the uniqueness of the representation (2.9), the coefficients c_{mj} corresponding to these roots are equal to zero. However, the LRF (2.10) governs a wider class of series than the minimal LRF (2.7).

Since the roots of the characteristic polynomial specify its coefficients uniquely, they also determine the corresponding LRF. Consequently, by removing the extraneous roots of the characteristic polynomial $P_r(\lambda)$, corresponding to the LRF (2.10), we can obtain the polynomial describing the minimal LRF of the series.

Example 2.1 *Annual seasonality*
Let the series F_N have the period 12 (for instance, this series describes a seasonality). Then it can be expressed as a sum of a constant and six harmonics:

$$f_n = c_0 + \sum_{k=1}^{5} c_k \cos(2\pi nk/12 + \phi_k) + c_6 \cos(\pi n). \tag{2.11}$$

Under the condition that $c_k \neq 0$ for $k = 0, \ldots, 6$ the series has finite-difference dimension 12. In other words, the characteristic polynomial of the minimal LRF governing the series (2.11) has 12 roots. All these roots have the modulus 1. Two real roots ($+1$ and -1) correspond to the first and the last terms in (2.11). The harmonic term with frequency $\omega_k = k/12$ generates two complex conjugate roots $\exp(\pm i 2\pi k/12)$, which have polar angles $\pm 2\pi k/12$.

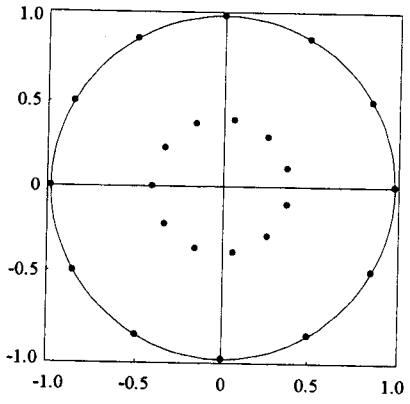

Figure 2.1 *Annual seasonality: main and extraneous roots.*

Let us now consider an LRF that is not minimal. Let N be large enough. If we select certain $L > 13$ and take $r = 12$, $\mathfrak{L}_r = \mathfrak{L}^{(L)}(F_N)$, then the vector

$\mathcal{R} = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ defined in (2.1) produces the LRF

$$f_{i+L-1} = a_1 f_{i+L-2} + \ldots + a_{L-1} f_i, \tag{2.12}$$

which is not minimal but governs the series (2.11).

Let us take $c_0 = \ldots = c_6 = 1$, $\phi_1 = \ldots = \phi_5 = 0$ and $L = 24$. The roots of the characteristic polynomial of the formula (2.12) are depicted in Fig. 2.1. We can see that the main 12 roots of the polynomial form a regular dodecagon, with the vertices on the unit circle of the complex plane. Eleven extraneous roots can be seen around zero; they have small moduli.

2.2.2 *Recurrent continuation of time series*

If the time series F_N is governed by an LRF (2.10) of dimension $r < N$, then there exists a natural *recurrent continuation* of such a series produced by the same formula (2.10). Whether LRF is minimal or not is of no importance since the extraneous roots have no influence on the series F_N.

(a) L-continuation

It is important to reformulate the concept of recurrent continuation in purely geometrical terms. Let us start with a definition.

Consider a time series $F_N = (f_0, \ldots, f_{N-1})$ and fix a window length $1 < L < N$. Denote by $X_1, \ldots, X_K$ the corresponding L-lagged vectors, and set $\mathfrak{L}^{(L)} = \mathrm{span}(X_1, \ldots, X_K)$. Let $d = \dim \mathfrak{L}^{(L)}$. (In other terms, the L-rank of the series F_N is equal to d.) Evidently, $d \le \min(L, K)$.

We say that the series F_N *admits a continuation in* $\mathfrak{L}^{(L)}$ (or, briefly, *admits L-continuation*) if there exists a uniquely defined $\widetilde{f}_N$ such that all L-lagged vectors of the series $\widetilde{F}_{N+1} = (f_0, \ldots, f_{N-1}, \widetilde{f}_N)$ belong to $\mathfrak{L}^{(L)}$. In this case, the series $\widetilde{F}_{N+1}$ (as well as the number $\widetilde{f}_N$) will be called the *one-step L-continuation* of the series F_N.

Theorem 5.4 and Remark 5.9 in Section 5.3 provide the complete description of those series that admit L-continuation. For the moment, the following is important.

1. If $e_L \in \mathfrak{L}^{(L)}$, then F_N does not admit L-continuation. As a consequence, if $d = L$, then the series cannot be L-continued since the uniqueness condition does not hold.
2. If $d < L \le K$ and $e_L \notin \mathfrak{L}^{(L)}$, then the series F_N admits L-continuation. From now on we assume that these assumptions concerning $\mathfrak{L}^{(L)}$ are satisfied.
3. The one-step L-continuation of the series F_N can be performed by the formula

$$\widetilde{f}_N = \sum_{k=1}^{L-1} a_k f_{N-k}, \tag{2.13}$$

where the vector $\mathcal{R} = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ is defined in the formula (2.1) applied to the space $\mathfrak{L}_r = \mathfrak{L}^{(L)}$.

4. The series F_N is governed by the same LRF (2.13), that is

$$f_{i+L} = \sum_{k=1}^{L-1} a_k f_{i+L-k}, \qquad 0 \le i \le N - L - 1.$$

5. If the series F_N admits a one-step L-continuation, then it can be L-continued for an arbitrary number of steps. Therefore, we can consider an infinite series F which is the L-continuation of F_N.
6. Let the series F_N satisfy an LRF

$$f_{i+d_0} = \sum_{k=1}^{d_0} b_k f_{i+d-k}, \qquad 0 \le i \le N - d_0 - 1, \tag{2.14}$$

and $d_0 \le \min(L-1, K)$. Then $d \le d_0$, $e_L \notin \mathfrak{L}^{(L)}$ and the series will admit L-continuation, which can be produced by the same formula (2.14).

These properties are not surprising in view of the results discussed above concerning the correspondence between the series with $\mathrm{fdim}(F_N) = d$ and $\mathrm{rank}_L(F_N) = d$. Reformulated in terms of continuation, this correspondence means that under the conditions $\mathrm{rank}_L(F_N) < L \le K$ and $e_L \notin \mathfrak{L}^{(L)}$ the concepts of recurrent continuation and L-continuation are equivalent.

(b) Recurrent continuation and Basic SSA forecasting

Let us return to the forecasting algorithm of Section 2.1, considering the case of Basic SSA R-forecasting.

Suppose that $\mathfrak{L}_r = \mathfrak{L}^{(L)}$, $e_L \notin \mathfrak{L}^{(L)}$ and $r < L \le K$. Then

$$r = \mathrm{rank}_L(F_N) = \mathrm{fdim}(F_N)$$

and the series F_N is governed by an LRF of order r. In other words, the series F_N admits L-continuation.

Since the vectors X_i belong to the linear space $\mathfrak{L}_r$, the matrix $\widetilde{\mathbf{X}}$ of the forecasting algorithm coincides with the trajectory matrix $\mathbf{X}$ for the initial series F_N.

Denote by F_{N+M} recurrent continuation of the series F_N for M steps. This continuation can be performed with the help of the LRF (2.6), as the latter governs the series F_N. By the algorithm description, the forecasting formula (2.2) is produced by the same LRF (2.6).

Therefore, the series G_{N+M} defined by the formula (2.2) is equal to F_{N+M} and the SSA R-forecasting algorithm with $\mathfrak{L}_r = \mathfrak{L}^{(L)}$ produces recurrent continuation of the series F_N. The vector form (2.4) of the algorithm corresponds to the L-continuation.

To obtain the vector $\mathcal{R}$, we must have an orthonormal basis of the linear space $\mathfrak{L}^{(L)}$, see formula (2.1). Dealing with SSA, the SVD of the trajectory matrix

$\mathbf{X}$ for the series F_N provides us with the eigenvectors (left singular vectors) $U_1, \ldots, U_r$, which form a natural basis of $\mathfrak{L}^{(L)}$. Therefore, if the series F_N admits L-continuation, the latter can be performed with the help of SSA.

Other choices of $\mathfrak{L}_r$ can lead to continuation of the series components. Let $F_N = F_N^{(1)} + F_N^{(2)}$ with nonzero $F_N^{(1)}$ and $F_N^{(2)}$. Denote the L-lagged vectors of the series $F_N^{(1)}$ by $X_1^{(1)}, \ldots, X_K^{(1)}$ and set $\mathfrak{L}^{(L,1)} = \operatorname{span}(X_1^{(1)}, \ldots, X_K^{(1)})$.

Let $r = \dim \mathfrak{L}^{(L,1)}$ and assume that $r < L \le K$ and $e_L \notin \mathfrak{L}^{(L,1)}$. Then the series $F_N^{(1)}$ admits L-continuation and the SSA R-forecasting algorithm with $\mathfrak{L}_r = \mathfrak{L}^{(L,1)}$ performs this continuation.

Suppose that series $F_N^{(1)}$ and $F_N^{(2)}$ are strongly separable for the window length $L \le K$ (see Section 1.5) and denote by $\mathbf{X}$ the trajectory matrix of the series F_N. Then the SVD of the matrix $\mathbf{X}$ produces both the space $\mathfrak{L}_r$ and the series $F_N^{(1)}$.

Indeed, let U_i $(i = 1, \ldots, L)$ be the eigenvectors of the matrix $\mathbf{S} = \mathbf{X}\mathbf{X}^{\mathrm{T}}$ and let $I = \{j_1, \ldots, j_r\} \subset \{1, \ldots, L\}$ be the set of indices corresponding to the time series $F_N^{(1)}$. If we take $P_i = U_{j_i}$, $i = 1, \ldots, r$, then $\mathfrak{L}_r = \operatorname{span}(P_1, \ldots, P_r)$ and $r < L$. The series $F_N^{(1)}$ can be obtained in terms of the resultant Hankel matrix, which is produced by the grouping of the elementary matrices corresponding to the set of indices I.

Therefore, Basic SSA gives rise to the continuation of the series component which is accomplished by the Basic SSA R-forecasting algorithm. Note that if $F_N^{(1)}$ and $F_N^{(2)}$ are strongly separable, then both dimensions of their trajectory spaces are smaller than L.

2.2.3 *Approximate continuation*

The problems of exact continuation have mainly a theoretical and methodological sense. In practice, it is not wise to assume that the series obtained by measurements is governed by some LRF of relatively small dimension. Thus, we pass to the concept of approximate continuation, which is of greater importance in practice.

(a) Approximate separability and forecasting errors

Let $F_N = F_N^{(1)} + F_N^{(2)}$ and suppose that the series $F_N^{(1)}$ admits a recurrent continuation. Denote by d the dimension of the minimal recurrent formula governing $F_N^{(1)}$. If $d < \min(L, K)$, then $d = \operatorname{rank}_L(F_N^{(1)})$.

If $F_N^{(1)}$ and $F_N^{(2)}$ are strongly separable for some window length L, then we can perform recurrent continuation of the series $F_N^{(1)}$ by the method described in Section 2.2.2. We now assume that $F_N^{(1)}$ and $F_N^{(2)}$ are approximately strongly separable and discuss the problem of approximate continuation of the series $F_N^{(1)}$.

If $F_N^{(2)}$ is small enough and signifies an error or noise, this continuation can be regarded as a forecast of the signal $F_N^{(1)}$ in the presence of noise $F_N^{(2)}$. In

other cases we can describe the problem as that of forecasting an interpretable component $F_N^{(1)}$ of F_N: for example, forecasting its trend or seasonal component.

As above, to do the continuation we use the Basic SSA R-forecasting algorithm described in Section 2.1. Formally, we assume that the following conditions hold.

1. The series of length N and window length L provide approximate strong separability of the series $F_N^{(1)}$ and $F_N^{(2)}$.
2. Let
$$\mathbf{X} = \sum_i \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$$
be the SVD of the trajectory matrix $\mathbf{X}$ of the series F_N. Then the choice of the eigentriples $\{(\sqrt{\lambda_i}, U_i, V_i)\}_{i \in I}$, $I = (i_1, \ldots, i_r)$, associated with $F_N^{(1)}$ allows us to achieve (approximate) separability.
3. $d \stackrel{\text{def}}{=} \operatorname{fdim}(F_N^{(1)}) \le r < L \le K$.
4. $e_L \notin \operatorname{span}(U_i, i \in I)$. In other terms, $\sum_{i \in I} u_{iL}^2 < 1$, where u_{iL} is the last component of the eigenvector U_i.

If these conditions hold, then we can apply the (Basic) SSA R-forecasting algorithm, taking $\mathfrak{L}_r = \operatorname{span}(U_i, i \in I)$ and $P_j = U_{i_j}$. The result $g_N, \ldots, g_{N+M-1}$ is called the *approximate recurrent continuation* of the series F_N.

Let us discuss the features of this forecasting method. The forecast series g_n ($n \ge N$) defined by (2.2), generally does not coincide with recurrent continuation of the series $F_N^{(1)}$. The errors have two origins. The main one is the difference between the linear space $\mathfrak{L}_r$ and $\mathfrak{L}^{(L,1)}$, the trajectory space of the series $F_N^{(1)}$. Since the LRF (2.2) is produced by the vector $\mathcal{R}$ and the latter is strongly related to the space $\mathfrak{L}_r$ (see Proposition 5.5 of Chapter 5), the discrepancy between $\mathfrak{L}_r$ and $\mathfrak{L}^{(L,1)}$ produces an error in the LRF governing the forecast series. In particular, the finite-difference dimension of the forecast series g_n ($n \ge N$) is generally greater than d.

The other origin of the forecasting errors lies in the initial data for the forecast. For recurrent continuation, the initial data is $f_{N-L+1}^{(1)}, \ldots, f_{N-1}^{(1)}$, where $f_n^{(1)}$ is the nth term of the series $F_N^{(1)}$. In the Basic SSA R-forecasting algorithm, the initial data consists of the last L–1 terms $g_{N-L+1}, \ldots, g_{N-1}$ of the reconstructed series. Since generally $f_n^{(1)} \ne g_n$, the initial data produces its own error of forecasting.

On the other hand, if the quality of approximate separability of $F_N^{(1)}$ and $F_N^{(2)}$ is rather good and we select the proper eigentriples associated with $F^{(1)}$, then we can expect that the linear spaces $\mathfrak{L}_r$ and $\mathfrak{L}^{(L,1)}$ are close. Therefore, the coefficients in the LRF (2.2) are expected to be close to those of the LRF governing recurrent continuation of the series $F_N^{(1)}$. Analogously, approximate separability implies that the reconstructed series g_n is close to $f_n^{(1)}$, and therefore the errors of the initial forecasting data are small. As a result, in this case we can expect that

the Basic SSA R-forecasting procedure provides a reasonable approximation to recurrent continuation of $F_N^{(1)}$, at least in the first few steps.

The following artificial example illustrates the role of separability in forecasting.

Example 2.2 *Separability and forecasting*
Let us consider the series $F_N = F_N^{(1)} + F_N^{(2)}$ with $N = 100$,

$$f_n = f_n^{(1)} + f_n^{(2)}, \quad f_n^{(1)} = 3a^n, \quad f_n^{(2)} = \sin(2\pi n/10)$$

and $a = 1.01$. Note that the series $F_N^{(2)}$ has finite-difference dimension 2 and $F_N^{(1)}$ is governed by the minimal LRF $f_n^{(1)} = af_{n-1}^{(1)}$.

If we want to forecast the series $F_N^{(1)}$, then we have to choose the window length L and take just one eigenvector of the corresponding SVD as the basis of the linear space $\mathfrak{L}_1$. (In this example, the leading eigenvector is acceptable for a wide range of L.)

Evidently, the forecasting result depends on L. The choice of the window length L can be expressed in terms of separability: a proper L ought to provide good separability characteristics. Let us compare the choice of two window lengths, $L = 50$ and $L = 15$, from the viewpoint of forecasting. Since exponential and harmonic series are asymptotically separable, the window length $L = 50$ seems to provide a good separation, while $L = 15$ should be regarded as too small.

The results for both Basic R-forecasting procedures are depicted in Fig. 2.2, where the top thick line starting at $n = 101$ corresponds to $L = 50$, and the analogous bottom thick line relates to $L = 15$. The thin line indicates the initial series F_N continued up to $n = 190$, which is the last forecasting point.

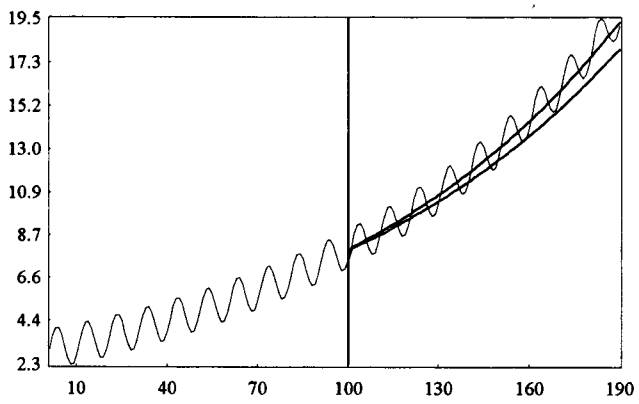

Figure 2.2 *Forecasting accuracy: two variants of window length.*

For $L = 50$, the choice of the first eigentriple in correspondence with $F_N^{(1)}$ leads to the w-correlation $\rho_{12}^{(w)} = 0.0001$ and the maximum cross-correlation $\rho^{(L,K)} = 0.034$. Therefore, the achieved separability should be regarded as rather good. If we take $L = 15$, then we obtain $\rho_{12}^{(w)} = 0.0067$ and $\rho^{(L,K)} = 0.317$, which means that the separation is poorer.

If we compare the forecasting results at the point $n = 190$, then we observe that the window length $L = 50$ provides the relative forecast error of about 2%, while the choice $L = 15$ gives almost 9%. This difference is not surprising since the window length $L = 15$ is too small for achieving good separability.

Note that both forecasts underestimate the true series. This can be explained in terms of the characteristic polynomials. Indeed, the main root of the polynomial $P_{14}(\lambda)$ corresponding to $L = 15$ is equal to 1.0091. The analogous root for $L = 50$ is 1.0098. The (single) root of the polynomial corresponding to the minimal LRF governing $F_N^{(1)}$ is $a = 1.01$. The arrangement of the roots coincides with the arrangement of the two forecasts and the exponential series $f_n^{(1)} = a^n$.

(b) Approximate continuation and the characteristic polynomials

Let us return to the errors of separability and forecasting. The discrepancies between $\mathfrak{L}_r$ and $\mathfrak{L}^{(L,1)}$ can be described in terms of the characteristic polynomials. We have three LRFs: (i) the minimal LRF of dimension d governing the series $F_N^{(1)}$, (ii) the continuation LRF of dimension $L-1$, which also governs $F_N^{(1)}$, but produces $L-d-1$ extraneous roots in its characteristic polynomial P_{L-1}, and (iii) the forecasting LRF governing the forecast series g_n $(n \geq N)$. The characteristic polynomial $P_{L-1}^{(f)}$ of the forecasting LRF also has $L-1$ roots.

If $\mathfrak{L}_r$ and $\mathfrak{L}^{(L,1)}$ are close, then the coefficients of continuation and forecasting recurrent formulae must be close too. Therefore, all simple roots of the forecasting characteristic polynomial $P_{L-1}^{(f)}$ must be close to that of the continuation polynomial P_{L-1}. The roots λ_m with multiplicities $k_m > 1$ could be perturbed in a more complex manner.

Example 2.3 *Perturbation of the multiple roots*
Let us consider the series F_N with

$$f_n = (A + 0.1\,n) + \sin(2\pi n/10), \quad n = 0, \ldots, 199.$$

Evidently, $F_N = F_N^{(1)} + F_N^{(2)}$ with the linear series $F_N^{(1)}$ defined by $f_n^{(1)} = A + 0.1\,n$ and the harmonic series $F_N^{(2)}$ corresponding to $f_n^{(2)} = \sin(2\pi n/10)$.

The series F_N has finite-difference dimension $\mathrm{fdim}(F_N) = 4$. Therefore, any LRF governing F_N produces a characteristic polynomial with four main roots. These main roots do not depend on A; the linear part of the series generates one real root $\lambda = 1$ of multiplicity 2, while the harmonic series corresponds to two complex conjugate roots with modulus $\rho = 1$ and $\omega = 0.1$.

Our aim is to forecast the series $F_N^{(1)}$ for $A = 0$ and $A = 50$ with the help of the Basic SSA forecasting algorithm. In both cases, we take the window length $L = 100$ and choose the eigentriples that correspond to the linear part of the initial time series F_N. (For $A = 0$ we take the two leading eigentriples, while for $A = 50$ the appropriate eigentriples have the ordinal numbers 1 and 4.) Since the series $F_N^{(1)}$ and $F_N^{(2)}$ are not exactly separable for any choice of L, we deal with approximate separability.

The forecasting polynomials $P_{L-1}^{(f)}$ with $A = 0$ and $A = 50$ demonstrate different splitting of the double root $\lambda = 1$ into two simple ones. For $A = 0$ there appear two complex conjugate roots with $\rho = 1.002$ and $\omega = 0.0008$, while in the case $A = 50$ we obtain two real roots equal to 1.001 and 0.997. All extraneous roots are less than 0.986.

This means that for $A = 0$ the linear series $F_N^{(1)}$ is approximated by a low-frequency harmonic with a slightly increasing exponential amplitude. In the case $A = 50$ the approximating series is the sum of two exponentials, one of them is slightly increasing and another one is slightly decreasing.

These discrepancies lead to quite different long-term forecasting results: oscillating for $A = 0$ and exponentially increasing for $A = 50$.

In the case of a large discrepancy between $\mathfrak{L}_r$ and $\mathfrak{L}^{(L,1)}$, both the main and the extraneous roots of the continuation polynomial can differ significantly, and the error of the forecasting procedure can be rather large.

Evidently, such an error depends on the order $L-1$ of the characteristic polynomials as well; the bigger the number of the perturbed extraneous roots, the less precise the forecasting procedure may become.

On the other hand, the conditions for approximate separability are usually asymptotic and require relatively large L. In practice, this means that we have to take the smallest window length L providing a sufficient (though approximate) separability.

2.3 Modifications to Basic SSA R-forecasting

The Basic SSA R-forecasting algorithm discussed in Section 2.2 should be regarded as the main forecasting algorithm due to its direct relation to the linear recurrent formulae. Nevertheless, there exist several natural modifications to this algorithm that can give better forecasts in specific situations.

2.3.1 SSA vector forecasting

Let us return to Basic SSA and assume that our aim is to extract a certain additive component $F_N^{(1)}$ from a series F_N. In this algorithm, for an appropriate window length L, we obtain the SVD of the trajectory matrix of the series F_N and select the eigentriples $(\sqrt{\lambda_i}, U_i, V_i)$, $i \in I = (j_1, \ldots, j_r)$, corresponding to $F_N^{(1)}$. Then

we obtain the resultant matrix

$$\mathbf{X}_I = \sum_{i \in I} \sqrt{\lambda_i} U_i V_i^{\mathrm{T}}$$

and, after diagonal averaging, we obtain the reconstructed series $\widetilde{F}_N^{(1)}$ that estimates $F_N^{(1)}$.

Note that the columns $\widehat{X}_1, \ldots, \widehat{X}_K$ of the resultant matrix $\mathbf{X}_I$ belong to the linear space $\mathfrak{L}_r = \mathrm{span}(U_i, i \in I)$. If $F_N^{(1)}$ is strongly separable from $F_N^{(2)} \stackrel{\mathrm{def}}{=} F_N - F_N^{(1)}$, then $\mathfrak{L}_r$ coincides with $\mathfrak{L}^{(L,1)}$ (the trajectory space of the series $F_N^{(1)}$) and $\mathbf{X}_I$ is a Hankel matrix (in this case $\mathbf{X}_I$ is the trajectory matrix of the series $F_N^{(1)}$). If $F_N^{(1)}$ and $F_N^{(2)}$ are approximately strongly separable, then $\mathfrak{L}_r$ is close to $\mathfrak{L}^{(L,1)}$ and $\mathbf{X}_I$ is approximately a Hankel matrix.

Briefly, the idea of 'vector forecasting' can be expressed as follows. Let us imagine that we can continue the sequence of vectors $\widehat{X}_1, \ldots, \widehat{X}_K$ for M steps in such a manner that:

1. The continuation vectors Z_m $(K < m \le K + M)$ belong to the same linear space $\mathfrak{L}_r$.
2. The matrix $\mathbf{X}_M = [\widehat{X}_1 : \ldots : \widehat{X}_K : Z_{K+1} : \ldots : Z_{K+M}]$ is approximately a Hankel matrix.

Having obtained the matrix $\mathbf{X}_M$ we can obtain the series G_{N+M} by diagonal averaging. Since the first elements of the reconstructed series $\widetilde{F}_N^{(1)}$ coincide with the elements of G_{N+M}, the latter can be considered to be a forecast of $F_N^{(1)}$.

Now let us give a formal description of the *SSA vector forecasting algorithm* (briefly, *V-forecasting*) in the same manner as was done in Section 2.1 for the SSA recurrent forecasting algorithm.

Preliminaries:

- The SSA vector forecasting algorithm has the same inputs and conditions as the SSA R-forecasting algorithm.
- The notation in (a)-(g) of Section 2.1 is kept. Let us introduce some more notation.

  Consider the matrix

$$\Pi = \mathbf{V}^{\nabla}(\mathbf{V}^{\nabla})^{\mathrm{T}} + (1 - \nu^2)\mathcal{R}\mathcal{R}^{\mathrm{T}}, \tag{2.15}$$

  where $\mathbf{V}^{\nabla} = [P_1^{\nabla} : \ldots : P_r^{\nabla}]$. The matrix Π is the matrix of the linear operator that performs the orthogonal projection $\mathbf{R}^{L-1} \mapsto \mathfrak{L}_r^{\nabla}$ (see Proposition 5.9 in Section 5.3), where $\mathfrak{L}_r^{\nabla} = \mathrm{span}(P_1^{\nabla}, \ldots, P_r^{\nabla})$.

  We define the linear operator $\mathcal{P}^{(v)} : \mathfrak{L}_r \mapsto \mathbf{R}^L$ by the formula

$$\mathcal{P}^{(v)}Y = \begin{pmatrix} \Pi Y_{\triangle} \\ \mathcal{R}^{\mathrm{T}} Y_{\triangle} \end{pmatrix}, \quad Y \in \mathfrak{L}_r. \tag{2.16}$$

SSA vector forecasting algorithm:

1. In the notation above we define the vectors Z_i as follows:

$$Z_i = \begin{cases} \widehat{X}_i & \text{for } i = 1, \ldots, K \\ \mathcal{P}^{(v)} Z_{i-1} & \text{for } i = K+1, \ldots, K+M+L-1. \end{cases} \tag{2.17}$$

2. By constructing the matrix $\mathbf{Z} = [Z_1 : \ldots : Z_{K+M+L-1}]$ and making its diagonal averaging we obtain a series $g_0, \ldots, g_{N+M+L-1}$.
3. The numbers $g_N, \ldots, g_{N+M-1}$ form the M terms of the SSA vector forecast.

If $\mathfrak{L}_r$ is spanned by certain eigenvectors obtained by Basic SSA, we shall call the corresponding algorithm the *Basic SSA vector forecasting algorithm*. Let us discuss its features.

(a) *Continuation*

If $\mathfrak{L}_r$ is the trajectory space of the series F_N (in other words, if we act under the assumptions of Section 2.2.2), then the result of the vector forecasting coincides with that of the recurrent one. Thus, in this case the V-forecasting algorithm performs recurrent continuation of the series F_N.

More precisely, in this situation the matrix Π is the identity matrix, and (2.16) coincides with (2.3). Furthermore, the matrix $\mathbf{Z}$ has Hankel structure and diagonal averaging is the identical operation.

The same coincidence holds if $F_N = F_N^{(1)} + F_N^{(2)}$, the series $F_N^{(1)}$ and $F_N^{(2)}$ are strongly separable, and $\mathfrak{L}_r$ is the trajectory space of the series $F_N^{(1)}$. The Basic SSA V-forecasting then performs recurrent continuation of $F_N^{(1)}$.

(b) *Forecasting*

Though the results are the same, the essentials of recurrent and vector forecasting are different. Briefly, recurrent forecasting performs recurrent continuation directly (with the help of LRF), while vector forecasting deals with L-continuation. In the case of approximate continuation, the two forecasting algorithms usually give different results.

In a typical situation, there is no time series such that the linear space $\mathfrak{L}_r$ (for $r < L-1$) is its trajectory space, and therefore (see Proposition 5.6) this space cannot be the trajectory space of the series to be forecasted. The recurrent forecasting method uses $\mathfrak{L}_r$ to obtain the LRF of the forecast series.

The vector forecasting procedure tries to perform the L-continuation of the series in $\mathfrak{L}_r$; any vector $Z_{i+1} = \mathcal{P}^{(v)} Z_i$ belongs to $\mathfrak{L}_r$, and Z_{i+1}^{∇} is as close to $(Z_i)_{\Delta}$ as it can be. The last component of Z_{i+1} is obtained from Z_{i+1}^{∇} by the LRF applied in the recurrent forecasting. Since the matrix $\mathbf{Z}$ is not a Hankel one, diagonal averaging works in the same manner as in Basic SSA.

(c) *Details*

Both forecasting methods have two general stages: diagonal averaging and continuation. For the recurrent forecasting, diagonal averaging is used to obtain the

reconstructed series, and continuation is performed by applying the LRF. In the vector forecasting method, these two stages are used in the reverse order; first, vector continuation in $\mathfrak{L}_r$ is performed and then diagonal averaging gives the forecast values.

Note that in order to get M forecast terms the vector forecasting procedure performs $M+L-1$ steps. The aim is the permanence of the forecast under variations in M: the M-step forecast ought to coincide with the first M values of the forecast for $M+1$ or more steps. In view of the features of diagonal averaging, we have to produce $L-1$ extra steps.

(d) *Comparison*
If the series admits recurrent continuation, then the results for both Basic SSA forecasting methods coincide. In the case of approximate continuation they differ. Typically, a poor approximation implies a large difference between the two forecasts.

In the case of approximate separability it is hard to compare the recurrent and vector forecasting methods theoretically. Generally, the approximate coincidence of the two forecasting results can be used as an argument in favour of the forecasting stability.

Recurrent forecasting is simpler to interpret due to the description of LRFs in terms of the characteristic polynomials. On the other hand, results of data analysis show that the vector forecasting method is usually more 'conservative' (or less 'radical') in those cases when the recurrent forecasting method demonstrates rapid increase or decrease.

2.3.2 *Toeplitz SSA forecasting*

Using Basic SSA recurrent and vector forecasting, we take $\mathfrak{L}_r$ to be spanned by certain eigenvectors U_k, $k \in I$, of the SVD applied to the trajectory matrix $\mathbf{X}$ of the series F_N. In other words, the basis vectors P_i of $\mathfrak{L}_r$ have the form $P_i = U_{j_i}$ (see Section 2.2.2). Other decompositions of the trajectory matrix lead to another choice of $\mathfrak{L}_r$.

If the original series can be regarded as a stationary one, then as defined in (1.34) the Toeplitz SSA decomposition

$$\mathbf{X} = \sum_{i=1}^{L} H_i Z_i^T$$

can be used in place of the SVD in Basic SSA. Here the H_i stands for the ith eigenvector of the Toeplitz lag-covariance matrix defined in (1.32). (See Section 1.7.2 in Chapter 1 for details.)

Let us consider the SSA R-forecasting algorithm of Section 2.1. If we select a set of indices $I = (j_1, \ldots, j_r)$ and take $P_i = H_{j_i}$ as the basis vectors in $\mathfrak{L}_r$, then

we obtain the *Toeplitz SSA R-forecasting algorithm*. Evidently, one can use the vector forecasting variant in Toeplitz forecasting as well.

As was mentioned in Section 1.7.2, for relatively short intervals of stationary-like series, the Toeplitz SSA may give better separability characteristics than Basic SSA. Therefore, if we have a problem of continuation of a sum of several harmonic components of a stationary series, then Toeplitz forecasting may have an advantage.

Moreover, if L is much smaller than $K = N - L + 1$, then the Toeplitz lag-covariance matrix has a more regular structure than the standard lag-covariance matrix used in Basic SSA. The eigenvectors of the Toeplitz lag-covariance matrix are also more regular. Since forecasting is based on the space $\mathfrak{L}_r$ generated by the eigenvectors (and does not use both the factor vectors and the singular values), for stationary time series Toeplitz SSA forecasting may give more stable results.

2.3.3 *Centring in SSA forecasting*

To elucidate the characteristics of the (single) centring variant of SSA forecasting, we start with a series that admits recurrent continuation.

Consider the series F_N with $\mathrm{fdim}(F_N) = d \geq 1$ and sufficiently large N. As was described in Section 2.2, if we take the window length L such that $d < \min(L, K)$ and suppose that the corresponding trajectory space $\mathfrak{L}^{(L)}$ is not a vertical one, then $\dim \mathfrak{L}^{(L)} = d$ and the choice $\mathfrak{L}_r = \mathfrak{L}^{(L)}$ leads to recurrent continuation of the series F_N, which is performed by SSA recurrent forecasting algorithm of Section 2.1. Let us consider another way of doing such a continuation.

By definition, the space $\mathfrak{L}^{(L)}$ is spanned by the L-lagged vectors $X_1, \ldots, X_K$ of the series F_N. In the same manner as in Section 1.7, we denote by $\mathcal{E} = \mathcal{E}_1(\mathbf{X})$ the vector of the row averages of the trajectory matrix $\mathbf{X}$. In other words, we set

$$\mathcal{E} = (X_1 + \ldots + X_K)/K. \tag{2.18}$$

Evidently, $\mathcal{E} \in \mathfrak{L}^{(L)}$. We set

$$\mathfrak{L}_{\mathcal{E}}^{(L)} = \mathrm{span}(X_1 - \mathcal{E}, \ldots, X_K - \mathcal{E}) = \mathfrak{L}^{(L)} - \mathcal{E}. \tag{2.19}$$

Then (see Section 4.4) the dimension $r \stackrel{\mathrm{def}}{=} \dim \mathfrak{L}_{\mathcal{E}}^{(L)}$ is equal to either d or $d-1$. Assume that $r \geq 1$ (the case $r = 0$ corresponds to a constant series F_N).

If $e_L \notin \mathfrak{L}_{\mathcal{E}}^{(L)}$, then according to the proof of Theorem 5.2, the last component y_L of any vector $Y \in \mathfrak{L}_{\mathcal{E}}^{(L)}$ is equal to the linear combination of its first $L-1$ components:

$$y_L = \sum_{k=1}^{L-1} a_k y_{L-k}, \tag{2.20}$$

where the vector $\mathcal{R} = (a_{L-1}, \ldots, a_1)^{\mathrm{T}}$ is obtained from $\mathfrak{L}_{\mathcal{E}}^{(L)}$ by the formula (2.1), with $P_1, \ldots, P_r$ standing for an orthonormal basis in $\mathfrak{L}_{\mathcal{E}}^{(L)}$.

Let us now consider the infinite series F, which is recurrent continuation of F_N, and denote by X_i $(i > K)$ the ith L-lagged vector in the series F. Since $X_1, \ldots, X_K$ span the space $\mathfrak{L}^{(L)}$, and $X_i \in \mathfrak{L}^{(L)}$ for any $i > K$, it follows that $X_i - \mathcal{E} \in \mathfrak{L}_{\mathcal{E}}^{(L)}$ for any i.

Let us denote by $z_k^{(i)}$ the kth component of the vector $X_i - \mathcal{E}$. In view of (2.20) we obtain

$$z_L^{(i)} = \sum_{k=1}^{L-1} a_k z_{L-k}^{(i)}. \tag{2.21}$$

Rewriting (2.21) in terms of $X_i = (f_{i-1}, \ldots, f_{i+L-2})^{\mathrm{T}}$ we come to the equalities

$$f_{i+L-2} = \sum_{k=1}^{L-1} a_k f_{i+L-2-k} + \varepsilon_L - \mathcal{R}^{\mathrm{T}} \mathcal{E}^{\nabla}, \quad i \geq 1, \tag{2.22}$$

where ε_L is the last component of the vector $\mathcal{E}$.

Thus, we have arrived at the *heterogeneous linear recurrent formula*, governing the series F_N and performing its recurrent continuation. Evidently, if $\mathcal{E} = \mathbf{0}_{\mathrm{L}}$, then (2.22) coincides with recurrent continuation formula which is obtained in terms of $\mathfrak{L}^{(L)}$, see Section 2.2.

The transition from the trajectory space $\mathfrak{L}^{(L)}$ to the space (2.19) is considered in Sections 4.4 and 1.7, where the features of the centring versions of the SVD and Basic SSA are discussed. In terms of these Sections, $\mathfrak{L}_{\mathcal{E}}^{(L)}$ corresponds to single centring.

Single centring ideas give rise to versions of both recurrent and vector SSA forecasting algorithms for Basic and Toeplitz forecasting. Let us describe these versions in the formal manner of Section 2.1. For brevity, we present only the modified items within the description of the algorithms.

There are two versions of these modifications. If we are reconstructing a component of a time series with the help of the centring variant of the Basic (or Toeplitz) SSA, we can either include the average triple into the list of the eigentriples selected for reconstruction or not. These two possibilities are kept in the *centring variant of SSA forecasting*.

Now let $\mathfrak{L}_r$ be a subspace of $\mathbf{R}^L$ of dimension $r < L$, $e_L \notin \mathfrak{L}_r$, and let $P_1, \ldots, P_r$ be some orthonormal basis of $\mathfrak{L}_r$.

If we do not take average triple for the reconstruction, then:

1. The matrix $\widehat{\mathbf{X}}$ (Section 2.1, *Notation and Comments*, item b) is defined as

$$\widehat{\mathbf{X}} = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} P_i P_i^{\mathrm{T}} (\mathbf{X} - \mathbf{A}), \tag{2.23}$$

   where $\mathbf{A} = [\mathcal{E} : \ldots : \mathcal{E}]$ and the vector $\mathcal{E}$ has the form (2.18).

2. Formula (2.2) and its vector version defined by (2.3) and (2.4) are kept for the recurrent variant of SSA centring forecasting. Analogously, for SSA vector forecasting, the formulae (2.16) and (2.17) are kept.

In the case when we take the average triple for the reconstruction, we have

1. Matrix $\widehat{\mathbf{X}}$ is defined as

$$\widehat{\mathbf{X}} = [\widehat{X}_1 : \ldots : \widehat{X}_K] = \sum_{i=1}^{r} P_i P_i^{\mathrm{T}} (\mathbf{X} - \mathbf{A}) + \mathbf{A},$$

in the same notation as (2.23).

2. (i) In recurrent forecasting, the formula (2.2) is modified as

$$g_i = \begin{cases} \widetilde{f}_i & \text{for } i = 0, \ldots, N-1, \\ \sum\limits_{j=1}^{L-1} a_j g_{i-j} + a & \text{for } i = N, \ldots, N+M-1 \end{cases}$$

with $a = \varepsilon_L - \mathcal{R}^{\mathrm{T}} \mathcal{E}^{\nabla}$. To modify its vector form (2.3), (2.4), we keep the latter formula and replace (2.3) by

$$\mathcal{P}^{(rc)} Y = \begin{pmatrix} Y_{\Delta} - \mathcal{E}^{\nabla} \\ \mathcal{R}^{\mathrm{T}} (Y_{\Delta} - \mathcal{E}^{\nabla}) \end{pmatrix} + \mathcal{E},$$

where the operator $\mathcal{P}^{(rc)}$ maps $\mathfrak{L}_r + \mathcal{E}$ to $\mathbf{R}^L$.

(ii) In SSA vector forecasting variant, the formula (2.17) is kept and (2.16) is replaced by

$$\mathcal{P}^{(vc)} Y = \begin{pmatrix} \Pi(Y_{\Delta} - \mathcal{E}^{\nabla}) \\ \mathcal{R}^{\mathrm{T}} (Y_{\Delta} - \mathcal{E}^{\nabla}) \end{pmatrix} + \mathcal{E}, \quad Y \in \mathfrak{L}_r + \mathcal{E}.$$

If we use *Basic SSA centring forecasting*, then the vectors P_i $(1 \le i \le r)$ are selected from the set of the SVD eigenvectors for the matrix $\mathbf{X} - \mathbf{A}$. In the *Toeplitz* variant, the Toeplitz decomposition of $\mathbf{X} - \mathbf{A}$ is used instead.

Note that the double centring variant of SVD (see Section 4.4) can hardly be used for forecasting in the style under consideration. The main reason for this is that the double centring is applied to both the rows and columns of the trajectory matrix, while the SSA forecasting algorithm of Section 2.1 and all its modifications and variants are based on the linear space $\mathfrak{L}_r$, which is associated only with the columns of the trajectory matrix.

2.3.4 *Other ways of modification*

There exist numerous versions of the forecasting methods based on the SSA ideas. Let us mention several of these versions, stating them as problems to be solved rather than as methods recommended for direct use in practice.

(a) Minimal recurrent formula: Schubert and reduction methods

The linear recurrent formula applied in the recurrent SSA forecasting algorithm has dimension $L-1$ (L is the window length), while the minimal recurrent formula governing the series F_N (if any) can have a much smaller dimension. There-

fore, for a window length sufficient for approximate separability, it is natural to look for the LRF of relatively small dimension to perform a reasonable forecast.

Assume that the series F_N admits a recurrent continuation. One way of finding its minimal LRF is described in Theorem 5.1 of Section 5.2, where such an LRF is explained in geometrical terms of the Schubert basis (*Schubert* method). Another possibility arises if we can distinguish the main and extraneous roots of the characteristic polynomial of the LRF. In this case we can remove the extraneous roots and come to the minimal formula (*reduction* method).

Both methods are theoretically exact if $\mathrm{fdim}(F_N) < \min(L, K)$. However, their practical usefullness is not at all obvious since we deal with approximate separability, which produces perturbations of all results.

The stability of the Schubert method under data perturbations has not yet been checked. Therefore, there is a danger that not only the coefficients of the obtained 'minimal' LRF but even its dimension can vary significantly under small variations in the data. Also, the method seems to be much more complicated than Basic SSA R-forecasting.

The modification of the Basic SSA R-forecasting algorithm based on the reduction of the polynomial roots works well if the main roots are properly indicated and the perturbation in the data is not very large. Otherwise the forecasting results can be unpredictable. An example of applying the reduction recurrent forecasting algorithm can be found in Section 2.6.1.

Note that both methods can be used only for recurrent forecasting. Moreover, the problem of the initial data arises again; the errors in the initial data for the minimal LRF can affect the forecast more severely than for an LRF of large dimension.

(b) The nearest subspace

If F_N admits recurrent continuation, then the choice $\mathfrak{L}_r = \mathfrak{L}^{(L)}$ leads to the LRF governing F_N. In the case of approximate separability, the forecasting LRF is calculated through the selected linear space $\mathfrak{L}_r$, which typically cannot be the trajectory space of any time series (see Proposition 5.6 in Section 5.2).

One can try to solve this annoying contradiction in the following manner. Let us state the problem of finding a linear space $\mathfrak{L}'_r$ as follows: (a) the space has the same dimension r as the initial space $\mathfrak{L}_r$, (b) $\mathfrak{L}'_r$ is the trajectory space of a certain time series, and (c) $\mathfrak{L}'_r$ is the closest to $\mathfrak{L}_r$ (the cosine of the angle between these spaces is maximum).

If the errors in data are not very large, then such a space can be regarded as an appropriate 'estimate' of the trajectory space of a series under recurrent continuation. The space $\mathfrak{L}'_r$ being found, the corresponding LRF of dimension $L-1$ appears, and the specific form of the forecast by this LRF depends on the initial data. Since the vector consisting of the last $L-1$ points of the reconstructed series does not generally belong to $\mathfrak{L}'_r$, we can perform its orthogonal projection onto this linear space and take the result for the initial forecast data.

We do not discuss here the general algorithmic problem of finding this nearest subspace. Let us consider the simplest case $r = 1$ when the space $\mathfrak{L}_1$ is spanned by a vector $X = (x_1, \ldots, x_L)^{\mathrm{T}}$, and $\mathfrak{L}_1'$ must be spanned by the vector $Y_a = (1, a, \ldots, a^{L-1})^{\mathrm{T}}$. Then the optimal a gives the maximum value for the expression $|(X, Y_a)|/||X||\,||Y_a||$ and can be obtained by simple calculations.

The optimal a being obtained, we must find the corresponding LRF, that is the vector $\mathcal{R}$ (see formula (2.1) in Section 2.1). In the one-dimensional case this problem is rather simple as well, since all the components of the formula (2.1) are expressed in terms of the single vector $P_1 = Y_a/||Y_a||$.

Omitting the calculations we present the result for the case $|a| \neq 1$:

$$\mathcal{R} = C(a)(1, a, \ldots, a^{L-2})^{\mathrm{T}} \tag{2.24}$$

with

$$C(a) = \frac{a^{L-1}(a^2 - 1)}{a^{2L-2} - 1}.$$

We can now apply the LRF so obtained to the appropriate initial data.

Evidently the one-dimensional case is convenient for the reduction of the extraneous polynomial roots; the LRF defined by (2.24) defines a characteristic polynomial with a single main root $\lambda = a$. Therefore, taking the last term of the reconstructed series as the initial point and applying the recurrent formula $f_n = af_{n-1}$, we make the forecast based on both ideas: that of the nearest subspace and the minimal LRF.

2.4 Forecast confidence bounds

According to the main SSA forecasting assumptions, the component $F_N^{(1)}$ of the series F_N ought to be governed by an LRF of relatively small dimension, and the residual series $F_N^{(2)} = F_N - F_N^{(1)}$ ought to be approximately strongly separable from $F_N^{(1)}$ for some window length L. In particular, $F_N^{(1)}$ is assumed to be a finite subseries of an infinite series $F^{(1)}$, which is a recurrent continuation of $F_N^{(1)}$. These assumptions cannot be ignored, but fortunately they hold for a wide class of practical problems.

To establish confidence bounds for the forecast, we have to apply even stronger assumptions, related not only to $F_N^{(1)}$, but to $F_N^{(2)}$ as well. First, let us consider $F_N^{(2)}$ as a finite subseries of an infinite random noise series $F^{(2)}$ that perturbs the signal $F^{(1)}$. The other assumptions can hardly be formulated in terms of $F_N^{(2)}$ only; they mainly deal with the residual series $\widetilde{F}_N^{(2)} = F_N - \widetilde{F}_N^{(1)}$, where $\widetilde{F}_N^{(1)}$ is the reconstructed component of F_N. Since $\widetilde{F}_N^{(1)} \approx F_N^{(1)}$, the features of $\widetilde{F}_N^{(2)}$ are strongly related to those of $F_N^{(2)}$. A more precise formulation of the additional assumptions depends on the problem we are solving and the method that we are applying.

Here we consider the following two problems, related to construction of the confidence bounds for the forecast. The first problem is to construct a confidence interval for the entire series $F = F^{(1)} + F^{(2)}$ at some future point in time $N+M$. The second problem can be formulated as a construction of confidence bounds for the signal $F^{(1)}$ at the same future point in time.

These two problems will be solved in different ways. The first uses the information about the forecast errors obtained by processing the series. This variant can be called the empirical one. The second requires additional information about the model governing the series $\widetilde{F}_N^{(2)}$ to accomplish a bootstrap simulation of the series F_N (see Efron and Tibshirani, 1986, Section 5, for general bootstrap concepts).

Let us briefly discuss both problems of constructing the confidence bounds for the Basic SSA R-forecasting method. All other SSA forecasting procedures can be treated analogously.

2.4.1 *Empirical confidence intervals for the forecast of the initial series*

Assume that we have already obtained the forecast value $\widetilde{f}_{N+M-1}^{(1)}$, that is, we have already performed M steps of the Basic SSA R-forecasting procedure. By definition, we use $\widetilde{f}_{N+M-1}^{(1)}$ as the forecast of the (future) term $f_{N+M-1}^{(1)}$ of the signal $F^{(1)}$. As was already mentioned, our problem is to build up a confidence interval for the (future) term f_{N+M-1} of the series F.

Let us consider the *multistart M-step recurrent continuation* procedure. We take a relatively small integer M and apply M steps of recurrent continuation produced by the forecasting LRF modifying the initial data from $(\widetilde{f}_0^{(1)}, \ldots, \widetilde{f}_{L-2}^{(1)})$ to $(\widetilde{f}_{K-M}^{(1)}, \ldots, \widetilde{f}_{N-M-1}^{(1)})$, $K = N - L + 1$.

The last points $g_{j+M+L-1}$ of these continuations can be compared with the values $f_{j+M+L-1}$ of the initial series F_N. We thus obtain the *multistart M-step residual series* H_{K-M+1} with

$$h_j^{(M)} = f_{j+M+L-2} - g_{j+M+L-2}, \quad j = 0, \ldots, K - M.$$

Suppose for the moment that the reconstructed series $\widetilde{F}_N^{(1)}$ coincides with $F_N^{(1)}$ and the forecasting LRF governs it. Then $g_k = f_k^{(1)}$ and the multistart M-step residual series coincides with the last $K - M + 1$ terms of the stationary noise series $F_N^{(2)}$.

If these suppositions are not valid, then $h_j^{(M)}$ does not coincide with $f_{j+M+L-2}^{(2)}$. Even so, let us assume that the multistart M-step residual series is stationary and ergodic in the sense that its empirical cumulative distribution function (c.d.f.) tends to the theoretical c.d.f. of the series as $N \to \infty$. Then, having the series H_{K-M+1} at hand, we can estimate certain of its quantiles (for example, the upper and lower 2.5% ones).

Note that the terms $g_{j+M+L-2}$ are obtained through the same number of steps with the same LRF as the forecast value $\widetilde{f}_{N+M-1}^{(1)}$, and their initial data is taken

from the same reconstructed series. Since forecasting requires the assumption that the series structure is kept in the future, the obtained empirical c.d.f. of the multistart M-step residual series can be used to construct the empirical confidence interval for f_{N+M-1}.

More formally, let us fix a confidence level γ ($0 < \gamma < 1$), and set $\alpha = 1-\gamma$. If $c^-_{\alpha/2}$ and $c^+_{\alpha/2}$ stand for the lower and upper $\alpha/2$-quantiles, calculated through the empirical c.d.f. of the multistart M-step residual series, then we obtain the *empirical confidence interval*

$$\left(\widetilde{f}^{(1)}_{N+M-1} + c^-_{\alpha/2},\ \widetilde{f}^{(1)}_{N+M-1} + c^+_{\alpha/2}\right),$$

which covers f_{N+M-1} with an approximate confidence level γ. Evidently, the number K has to be sufficiently large for the empirical c.d.f. to be stable.

If the multistart M-step residual series can be regarded as white noise, then the other variant of empirical confidence intervals is meaningful. Assuming the Gaussian white noise hypothesis, the standard symmetrical confidence bounds of f_{N+M-1} can be constructed with the help of the sample average and the sample variance of the multistart M-step residual series. Of course, the white noise hypothesis can be checked with the help of the standard statistical procedures.

2.4.2 *Bootstrap confidence bounds for the forecast of a signal*

Let us consider a method of constructing confidence bounds for the signal $F^{(1)}$ at the moment of time $N + M - 1$. In the unrealistic situation, when we know both the signal $F^{(1)}$ and the true model of the noise $F^{(2)}_N$, the Monte Carlo simulation can be applied to check the statistical properties of the forecast value $\widetilde{f}^{(1)}_{N+M-1}$ relative to the actual term $f^{(1)}_{N+M-1}$.

Indeed, assuming that the rules for the eigentriple selection are fixed, we can simulate S independent copies $F^{(2)}_{N,i}$ of the process $F^{(2)}_N$ and apply the forecasting procedure to S independent time series $F_{N,i} \stackrel{\text{def}}{=} F^{(1)}_N + F^{(2)}_{N,i}$. Then the forecasting results will form a sample $\widetilde{f}^{(1)}_{N+M-1,i}$ ($1 \le i \le S$), which should be compared against $f^{(1)}_{N+M-1}$. In this way the *Monte Carlo confidence bounds* for the forecast can be build up.

Since in practice we do not know the signal $F^{(1)}_N$, we cannot apply this procedure. Let us describe the bootstrap (for example, Efron and Tibshirani, 1986, Section 5) variant of the simulation for constructing the confidence bounds for the forecast.

Under a suitable choice of the window length L and the corresponding eigentriples, we have the representation $F_N = \widetilde{F}^{(1)}_N + \widetilde{F}^{(2)}_N$, where $\widetilde{F}^{(1)}_N$ (the reconstructed series) approximates $F^{(1)}_N$, and $\widetilde{F}^{(2)}_N$ is the residual series. Suppose now that we have a (stochastic) model of the residuals $\widetilde{F}^{(2)}_N$. (For instance, we can pos-

tulate some model for $F_N^{(2)}$ and, since $\widetilde{F}_N^{(1)} \approx F_N^{(1)}$, apply the same model for $\widetilde{F}_N^{(2)}$ with the estimated parameters.)

Then, simulating S independent copies $\widetilde{F}_{N,i}^{(2)}$ of the series $F_N^{(2)}$, we obtain S series $F_{N,i} \stackrel{\text{def}}{=} \widetilde{F}_N^{(1)} + \widetilde{F}_{N,i}^{(2)}$ and produce S forecasting results $\widetilde{f}_{N+M-1,i}^{(1)}$ in the same manner as in the Monte Carlo simulation variant.

More precisely, any time series $F_{N,i}$ produces its own reconstructed series $\widetilde{F}_{N,i}^{(1)}$ and its own forecasting linear recurrent formula LRF_i for the same window length L and the same set of the eigentriples. Starting at the last $L-1$ terms of the series $\widetilde{F}_{N,i}^{(1)}$, we perform M steps of forecasting with the help of its LRF_i to obtain $\widetilde{f}_{N+M-1,i}^{(1)}$.

As soon as the sample $\widetilde{f}_{N+M-1,i}^{(1)}$ $(1 \le i \le S)$ of the forecasting results is obtained, we can calculate its (empirical) lower and upper quantiles of a fixed level γ and obtain the corresponding confidence interval for the forecast. This interval (called the *bootstrap confidence interval*) can be compared with the forecast value $\widetilde{f}_{N+M-1}^{(1)}$ obtained from the initial forecasting procedure. A discrepancy between this value and the obtained confidence interval can be caused by the inaccuracy of the stochastic model for $\widetilde{F}_N^{(2)}$.

The average of the bootstrap forecast sample (*bootstrap average forecast*) estimates the mean value of the forecast, while the mean square deviation of the sample shows the accuracy of the estimate.

The simplest model for $\widetilde{F}_N^{(2)}$ is the model of Gaussian white noise. The corresponding hypothesis can be checked with the help of the standard tests for randomness and normality.

2.4.3 *Confidence intervals: comparison of forecasting variants*

The aim of this section is to compare different SSA forecasting procedures using several artificial series and Monte Carlo confidence intervals.

Let $F_N = F_N^{(1)} + F_N^{(2)}$, where $F_N^{(2)}$ is Gaussian white noise with standard deviation σ. Assume also that the signal $F_N^{(1)}$ admits a recurrent continuation. We can and shall perform a forecast of the series $F_N^{(1)}$ for M steps using different variants of SSA forecasting and appropriate eigentriples associated with $F_N^{(1)}$.

If the signal $F_N^{(1)}$ and its recurrent continuation are known, then we can apply the Monte Carlo procedure described in the previous section to check the accuracy of the forecasting results and compare different ways of forecasting.

To do that, we simulate a large number of independent copies $F_{N,i}^{(2)}$ of $F_N^{(2)}$, produce the time series $F_{N,i} = F_N^{(1)} + F_{N,i}^{(2)}$, and forecast their signal component $F_N^{(1)}$ using the eigentriples of the same ordinal numbers as that for the initial series F_N. Evidently this procedure is meaningful only if the choice of the eigentriples is stable enough for different realizations of the white noise $F_N^{(2)}$.

Monte Carlo forecast of the signal $F_N^{(1)}$ is useful in at least two respects: its average (called the *Monte Carlo average forecast*) shows the bias produced by the corresponding forecasting procedure, while the upper and lower quantiles indicate the role of the random component in the forecasting error.

Several effects will be illustrated with the help of this technique. First, we shall compare some forecasting variants from the viewpoint of their accuracy. The second matter to be demonstrated is the role of the proper window length. Lastly, we compare different variants of the confidence intervals in forecasting.

Throughout all the examples, we use the following notation: N stands for the length of the initial series, M is the number of forecasting steps, and σ denotes the standard deviation of the Gaussian white noise $F_N^{(2)}$. The confidence intervals are obtained in terms of the 2.5% upper and lower quantiles of the corresponding empirical c.d.f. using the sample size $S = 1000$.

(a) Periodic signal: recurrent and vector forecasting

Let $N = 100$, $M = 100$, $\sigma = 0.5$. Let us consider a periodic signal $F_N^{(1)}$ of the form

$$f_n^{(1)} = \sin(2\pi n/17) + 0.5\sin(2\pi n/10).$$

The series $F_N^{(1)}$ has difference dimension 4, and we use four leading eigentriples for its forecasting under the choice $L = 50$. The initial series $F_N = F_N^{(1)} + F_N^{(2)}$ and the signal $F_N^{(1)}$ (the thick line) are depicted in Fig. 2.3.

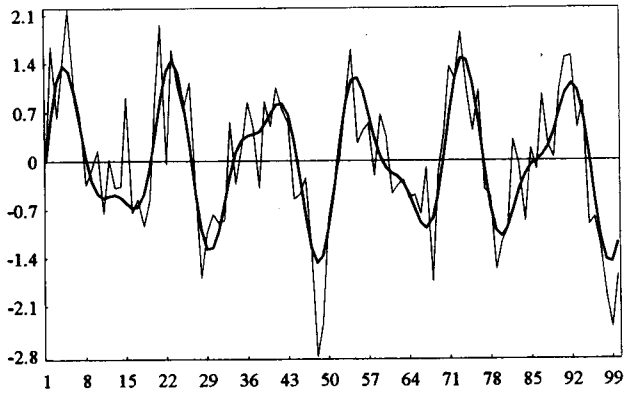

Figure 2.3 *Periodic signal and the initial series.*

Let us apply the Monte Carlo simulation for the Basic SSA recurrent and vector forecasting algorithms.

Fig. 2.4 shows the confidence Monte Carlo intervals for both methods and the true continuation of the signal $F_N^{(1)}$ (thick line). Confidence intervals for R-forecasting are marked by dots, while thin solid lines correspond to vector forecasting. We can see that these intervals practically coincide for relatively small numbers of forecasting steps, while the vector method has an advantage in the long-term forecasting.

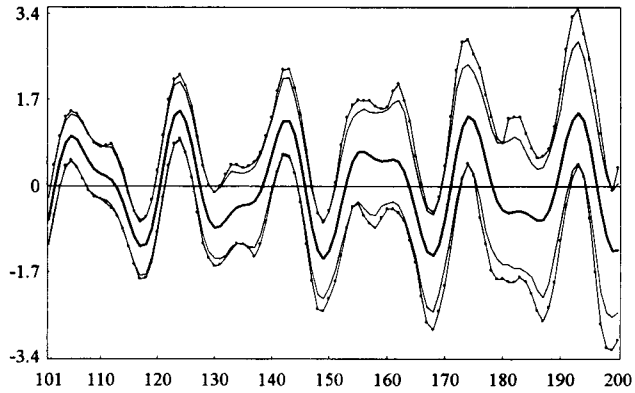

Figure 2.4 *Periodic signal: confidence intervals for the recurrent and vector forecasts.*

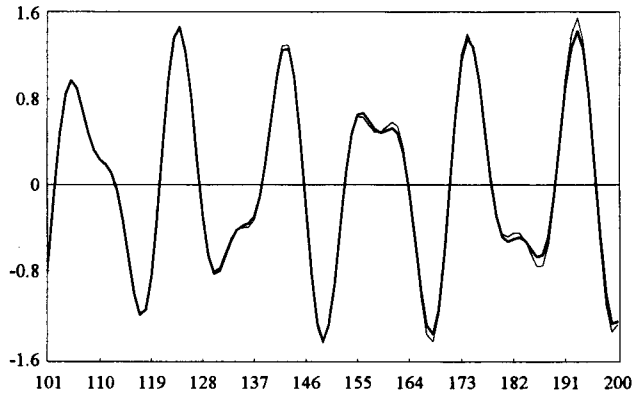

Figure 2.5 *Periodic signal: Basic Monte Carlo average R-forecast.*

The bias in the Basic SSA R-forecast is demonstrated in Fig. 2.5, where the thick line depicts the true continuation of the series $F_N^{(1)}$ and the thin line corresponds to the average of the Monte Carlo average R-forecast. We see that the bias is sufficiently small.

Note that the bias in the vector method almost coincides with that in the recurrent one. Therefore, the advantage of vector forecasting can be expressed mainly in terms of its stability rather than in the bias. The bias in both methods is caused by the nonlinear structure of the forecasting procedures.

(b) Periodic signal: Basic and Toeplitz recurrent forecasting

The same series with the same forecasting parameters serves as an example for comparing the Basic and Toeplitz R-forecasting methods. As usual, we apply the centring variant of the Toeplitz forecasting algorithm, though the results of the comparison do not depend on this choice.

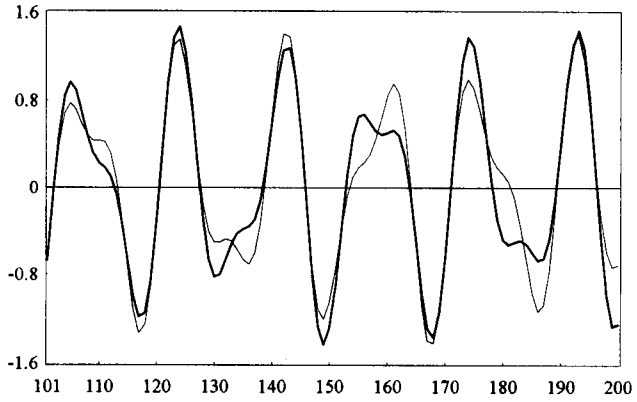

Figure 2.6 *Periodic signal: Toeplitz Monte Carlo average R-forecast.*

Fig. 2.6 is analogous to Fig. 2.5 and shows the bias in Toeplitz R-forecasting. In comparison with the Basic R-forecast, we see that the bias is rather large. The explanation lies in the fact that in contrast to Basic SSA, the four leading eigentriples in the Toeplitz SSA decomposition of the signal $F_N^{(1)}$ do not describe the entire signal; their share is approximately 99.8%. From the formal viewpoint, the Toeplitz decomposition of the trajectory matrix is not the minimal one (see Sections 4.2.1 and 1.7.2).

Indeed, if we consider the signal $F_N^{(1)}$ as the initial series and produce its Toeplitz forecast with $L = 50$ and 4 leading eigentriples, then the result will be very close to the Monte Carlo average forecast, presented in Fig. 2.6 (thin line, the thick line depicts the continuation of the series $F_N^{(1)}$).

The situation with the confidence intervals is different, see Fig. 2.7. The Monte Carlo confidence intervals for the Toeplitz forecast (depicted by thick lines) are typically inside that for the Basic forecast (thin lines). This is not surprising since the Toeplitz SSA gives more stable harmonic-like eigenvectors for stationary time series.

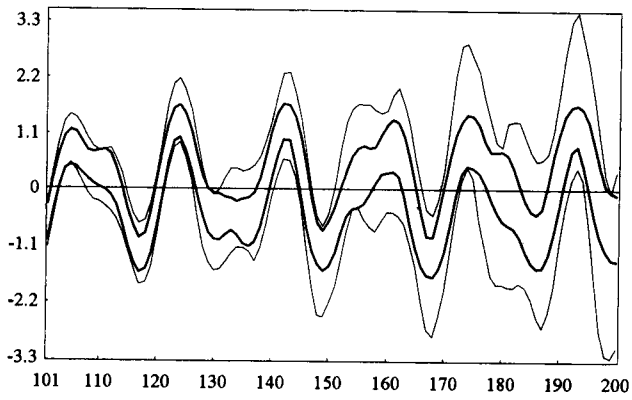

Figure 2.7 *Periodic signal: confidence intervals for the Basic and Toeplitz R-forecasts.*

Note that the confidence intervals for the Basic and Toeplitz forecasting algorithms are shifted relative to each other due to a large bias in the Toeplitz method. We conclude that Toeplitz forecasting proves to be less precise (on average), but more stable.

(c) Separability and forecasting

Consider the series $F_N^{(1)}$ with

$$f_n^{(1)} = 3a^n + \sin(2\pi n/10), \quad a = 1.01,$$

and $N = 100$. This series is governed by an LRF of dimension 3.

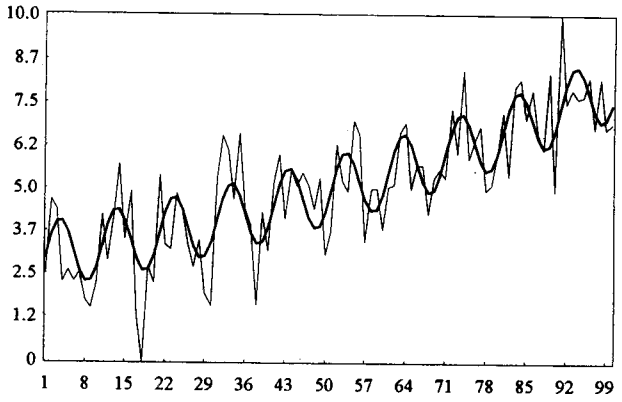

Figure 2.8 *Separability and forecasting: the signal and the initial series.*

Taking $\sigma = 1$ and two window lengths $L = 15$ and $L = 50$, we consider Basic SSA R-forecasting of the series $F_N = F_N^{(1)} + F_N^{(2)}$ for 90 steps. Our aim is to compare the accuracy of these two variants of forecasting of the signal $F_N^{(1)}$ with the help of the Monte Carlo simulation. The first three eigentriples are chosen for the reconstruction in both variants. The series F_N and the signal $F_N^{(1)}$ (thick line) are depicted in Fig. 2.8.

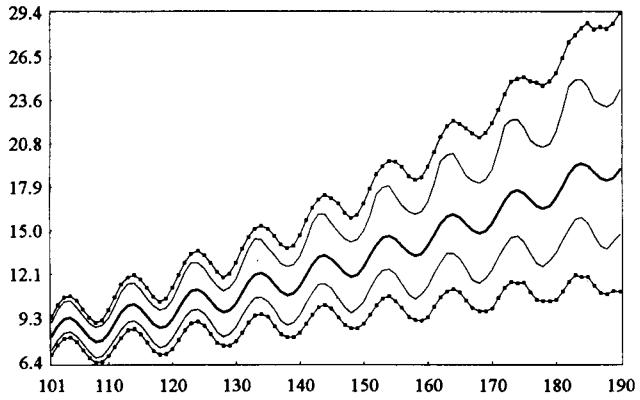

Figure 2.9 *Separability and forecasting: two confidence intervals.*

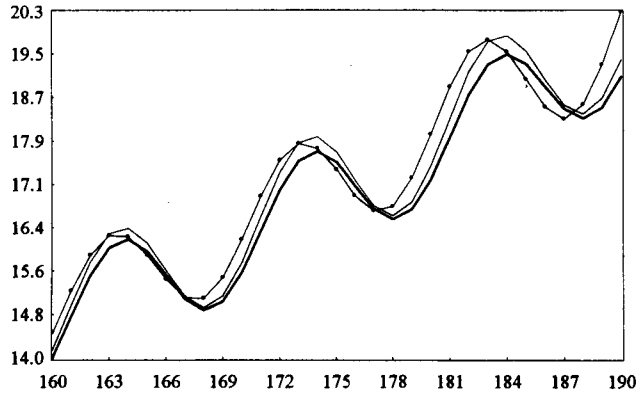

Figure 2.10 *Separability and forecasting: comparison of biases.*

The influence of separability on forecasting in the absence of noise has already been discussed (see Example 2.2 in Section 2.2). We now explain this influence in statistical terms of bias and confidence intervals.

Fig. 2.9 shows that the Monte Carlo forecasting confidence intervals for $L = 15$ (thin line marked with dots) are much wider than that for $L = 50$. This is not surprising since the separability characteristics for $L = 15$ are: $\rho_{12}^{(w)} = 0.0083$ and $\rho^{(L,K)} = 0.26$, while for $L = 50$ we have $\rho_{12}^{(w)} = 0.0016$ and $\rho^{(L,K)} = 0.08$.

Note that both confidence intervals are almost symmetric with respect to the true continuation of the signal (the thick line in Fig. 2.9). This means that in this example the choice of the window length does not have a big influence on the bias of the forecasts. Yet if we consider the last forecast points (Fig. 2.10), we can see that the choice $L = 50$ is again better. Indeed, the Monte Carlo average forecast for $L = 15$ (thin line, marked with dots) has a small but apparent phase shift relative to the true continuation (thick line), while for the choice $L = 50$ (thin line) there is almost no phase shift.

(d) Confidence intervals of different kinds

According to the discussion at the beginning of this section, we can construct three kinds of confidence interval for forecasting (see Section 2.4 for their detailed description).

First, as we know the true form of both the signal $F_N^{(1)}$ and the noise $F_N^{(2)}$, we can build the Monte Carlo confidence intervals, which can be considered to be the true confidence intervals for the signal forecast.

Second, we can apply the bootstrap simulation for the same purpose. Here we use the same Gaussian white noise assumption but calculate its variance in terms of the residuals of the reconstruction.

Third, the empirical confidence bounds for the forecast of the entire series $F_N = F_N^{(1)} + F_N^{(2)}$ can be built as well.

The last two methods are more important in practice since neither $F_N^{(1)}$ nor $F_N^{(2)}$ is usually known. Our aim is to compare three kinds of confidence bounds by a simple example.

Consider the exponential series $F_N^{(1)}$ with $f_n^{(1)} = 3a^n$, $a = 1.01$ and $N = 190$. As above, we assume that $F_N^{(2)}$ is a realization of the Gaussian white noise and take $\sigma = 1$. Since we want to deal with the empirical confidence intervals, we truncate the series at $n = 160$ and use the truncated series as the initial one. A comparison of the confidence intervals is performed for 30 Basic SSA R-forecasting steps with $L = 50$. Since $F_N^{(1)}$ is governed by an LRF of dimension 1, we take one leading eigentriple for reconstruction and forecasting in all cases.

The series F_N (thin oscillating line) is depicted in Fig. 2.11 together with its reconstruction, the Basic SSA R-forecast (thick lines) and the corresponding empirical intervals. The vertical line corresponds to the truncation point.

Figs. 2.12-2.14 show three variants of the confidence intervals on the background of the series F_N. Fig 2.12 represents the empirical intervals around the forecast of the signal $F_N^{(1)}$ (thick line). Since the empirical intervals are built for the entire series F_N, it is not surprising that they cover the series values. Note

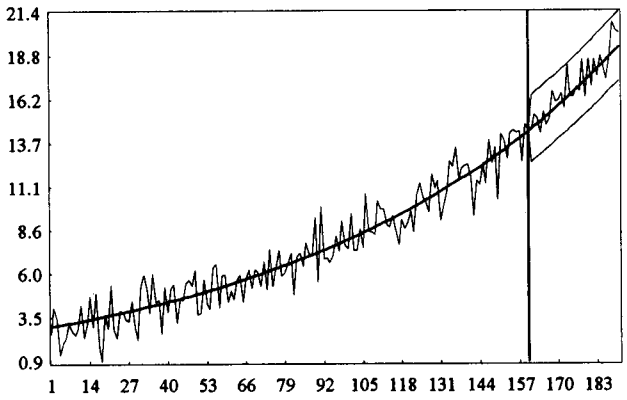

Figure 2.11 *Exponential signal: the initial series and forecast.*

that the length of the empirical confidence intervals is almost constant due to the homogeneity of the residuals used for their construction (see Section 2.4).

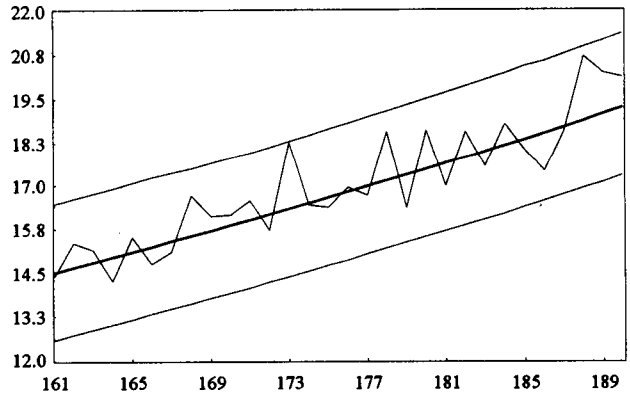

Figure 2.12 *Exponential signal: empirical confidence intervals.*

The bootstrap confidence intervals are shown in Fig. 2.13, where the thick line corresponds to the exponential signal $F_N^{(1)}$. The intervals are shifted relative to the signal (and they are symmetric relative to its forecast) because the bootstrap simulation uses the reconstructed series, which differs from the signal itself. Note that the empirical confidence intervals in Fig. 2.12 are also shifted relative to the signal $F_N^{(1)}$.

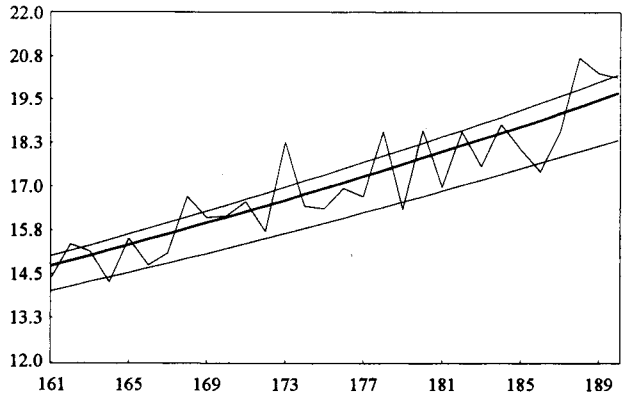

Figure 2.13 *Exponential signal: bootstrap confidence intervals.*

Lastly, the Monte Carlo confidence intervals are depicted in Fig. 2.14 together with the signal $F_N^{(1)}$ (thick line). In this case the intervals appear to be symmetric around the signal.

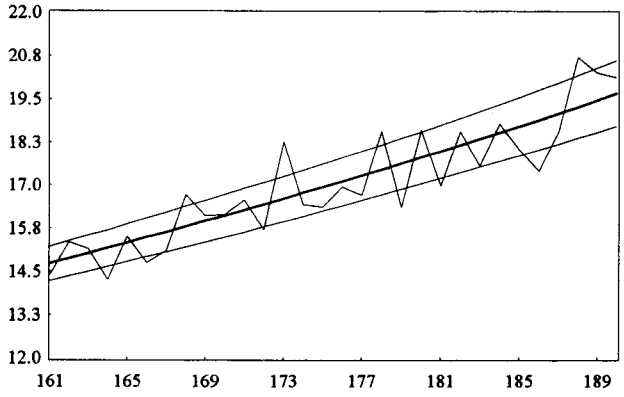

Figure 2.14 *Exponential signal: Monte Carlo confidence intervals.*

Comparing the intervals, we can see that the lengths of the bootstrap and Monte Carlo intervals are very similar and are smaller than those of the empirical intervals. The latter is natural since the first two bound the signal and the third one bounds the entire series.

One more difference is that the intervals obtained by simulation are enlarging in time, while the empirical ones are rather stable. Thus, we can use the empirical confidence intervals only for relatively short-term forecasting.

2.5 Summary and recommendations

Let us summarize the material of the previous sections, taking as an example the Basic SSA R-forecasting method. Other versions of SSA forecasting can be described and commented on similarly.

1. *Statement of the problem*
   We have a series $F_N = F_N^{(1)} + F_N^{(2)}$ and have the problem of forecasting its component $F_N^{(1)}$. If $F_N^{(2)}$ can be regarded as noise, then the problem is that of forecasting the signal $F_N^{(1)}$ in the presence of a noise $F_N^{(2)}$.
2. *The main assumptions*
   - The series $F_N^{(1)}$ admits a recurrent continuation with the help of an LRF of relatively small dimension d.
   - There exists a number L such that the series $F_N^{(1)}$ and $F_N^{(2)}$ are approximately strongly separable for the window length L. This is an important assumption since any time series $F_N^{(1)}$ is an additive component of F_N in the sense that $F_N = F_N^{(1)} + F_N^{(2)}$ with $F_N^{(2)} = F_N - F_N^{(1)}$. The assumption of (approximate) separability means that $F_N^{(1)}$ is a natural additive component of F_N from the viewpoint of the SSA method.
3. *Proper choice of parameters*
   Since we have to select the window length L providing a sufficient quality of separability and to find the eigentriples corresponding to $F_N^{(1)}$, all the major rules of Basic SSA are applicable here. Note that in this case we must separate $F_N^{(1)}$ from $F_N^{(2)}$, but we do not need the decomposition of the entire series $F_N = F_N^{(1)} + F_N^{(2)}$.
4. *Specifics and dangers*
   The SSA forecasting problem has some specifics in comparison with the Basic SSA reconstruction problem:
   - Since the chosen window length L produces an LRF of dimension $L-1$, which is applied as a recurrent continuation formula, the problem of extraneous roots for its characteristic polynomial becomes important. The choice $L = d+1$ with d standing for the dimension of the minimal LRF, must be optimal. Unfortunately, in practice, small values of L do not usually provide sufficient separability. As a result, one has to try to select the minimal window length that is greater than d and provides reasonable separability.
   - The linear space $\mathfrak{L}_r$ of dimension r determining the forecasting LRF is spanned by the eigenvectors of the chosen eigentriples. Since the condition $r \geq d$ has to be fulfilled, the number of eigentriples selected as corresponding to $F_N^{(1)}$ has to be at least d.
   - In Basic SSA, if we enlarge the set of proper eigentriples by some extra eigentriples with small singular values, then the result of reconstruction will

essentially be the same. When dealing with forecasting, such an operation can produce large perturbations since the space $\mathfrak{L}_r$ will be perturbed a lot; its dimension will be enlarged, and therefore the LRF governing the forecast will be modified. (Note that the magnitude of the extra singular values is not important in this case.) Hence, the eigentriples describing $F_N^{(1)}$ have to be determined very carefully.

5. *Characteristics of forecasting*
Let us mention several characteristics that might be helpful in judging the forecasting quality.

- *Separability characteristics.* All the separability characteristics considered in detail in Section 1.5 are of importance for forecasting.
- *Polynomial roots.* The roots of the characteristic polynomial of the forecasting LRF can give insight into the behaviour of the forecast. These polynomial roots can be useful in answering the following two questions:

  (a) We expect that the forecast has some particular form (for example, we expect it to be increasing). Do the polynomial roots describe such a possibility? For instance, an exponential growth has to be indicated by a single real root (slightly) greater than 1; if we try to forecast the annual seasonality, then pairs of complex roots with frequencies $\approx k/12$ have to exist, and so on.

  (b) Is it possible to obtain a hazard inconsistent forecast? In terms of the polynomial roots, each extraneous root increases such a possibility. Even so, if the modulus of the root is essentially less than 1, then a slight perturbation of the proper initial data should not produce large long-term errors. Since the polynomial roots with moduli greater than 1 correspond to the series components with increasing envelopes (see Section 2.2.1), large extraneous roots may cause problems even in short-term forecasting.

- *Verticality coefficient.* The verticality coefficient ν^2 is the squared cosine of the angle between the space $\mathfrak{L}_r$ and the vector e_L. The condition $\nu^2 < 1$ is necessary for forecasting. If ν^2 is close to 1, then, in view of (2.1), the coefficients of the forecasting LRF will be large and therefore some roots of the characteristic polynomial will have large moduli too. If the expected behaviour of the forecast does not suggest a rapid increase or decrease, then a large value of the verticality coefficient indicates a possible difficulty with the forecast. This typically means that extra eigentriples are taken to describe $F_N^{(1)}$ (alternatively, the approach in general is inappropriate).

6. *The role of the initial data*
Apart from the number M of forecast steps, the formal parameters of the Basic SSA R-forecasting algorithm are the window length L and the set I of eigentriples describing $F_N^{(1)}$. These parameters determine both the forecasting

LRF (2.1) and the initial data for the forecast. Evidently, the forecasting result essentially depends on this data, especially when the forecasting LRF has extraneous roots.

The SSA R-forecasting method uses the last terms $\widetilde{f}_{N-L+1}^{(1)}, \ldots, \widetilde{f}_{N-1}^{(1)}$ of the reconstructed series $\widetilde{F}_N^{(1)}$ as the initial forecasting data. Due to the properties of diagonal averaging, the last (and the first) terms of the series $F_N^{(1)}$ are usually reconstructed with a poorer precision than the middle ones. This effect may cause essential forecast errors.

For example, any linear (and nonconstant) series $f_n = an + b$ is governed by the minimal LRF $f_n = 2f_{n-1} - f_{n-2}$, which does not depend on a and b. The parameters a and b used in the forecast are completely determined by the initial data f_0 and f_1. Evidently, errors in this data may essentially modify the behaviour of the forecast (for example, change a tendency to increase into a tendency to decrease).

Thus, it is important to check the last points of the reconstructed series (for example, to compare them with the expected future behaviour of the series $F_N^{(1)}$).

7. *Reconstructed series and LRFs*
In the situation of strong separability of $F_N^{(1)}$ and $F_N^{(2)}$ and proper eigentriple selection, the reconstructed series is governed by the LRF which completely corresponds to the series $F_N^{(1)}$. Discrepancies in such a correspondence indicate possible errors: insufficient separability (which can be caused by the bad quality of the forecasting parameters) or general inefficiency of the model. Two characteristics of the correspondence may be useful here.

- *Global discrepancies.* Rather than using an LRF for forecasting, we can use it for approximation of either the whole reconstructed series or its subseries. For instance, if we take the first terms of the reconstructed series as the initial data (instead of the last ones) and make $N - L + 1$ steps of the procedure, we can check whether the reconstructed series can be globally approximated with the help of the LRF.
  Evidently, we can use another part of the reconstructed series as the initial data while taking into consideration the poor quality of its first terms or possible heterogeneity of the dynamics of the series $F_N^{(1)}$.
- *Local discrepancies.* The procedure above corresponds to long-term forecasting. To check the short-term correspondence of the reconstructed series and the forecasting LRF, one can apply a slightly different method.
  This method is used in Section 2.4.1 to construct empirical confidence intervals and is called the multistart recurrent continuation. According to it, for a relatively small Q we perform Q steps of the multistart recurrent continuation procedure, modifying the initial data from $(\widetilde{f}_0^{(1)}, \ldots, \widetilde{f}_{L-1}^{(1)})$ to $(\widetilde{f}_{K-Q}^{(1)}, \ldots, \widetilde{f}_{N-Q}^{(1)})$, $K = N - L + 1$. The continuation is computed