Estimation of parameters and eigenmodes of multivariate autoregressive models

ARNOLD NEUMAIER
Universität Wien
and
TAPIO SCHNEIDER
New York University

Dynamical characteristics of a complex system can often be inferred from analyses of a stochastic time series model fitted to observations of the system. Oscillations in geophysical systems, for example, are sometimes characterized by principal oscillation patterns, eigenmodes of estimated autoregressive (AR) models of first order. This paper describes the estimation of eigenmodes of AR models of arbitrary order. AR processes of any order can be decomposed into eigenmodes with characteristic oscillation periods, damping times, and excitations. Estimated eigenmodes and confidence intervals for the eigenmodes and their oscillation periods and damping times can be computed from estimated model parameters. As a computationally efficient method of estimating the parameters of AR models from high-dimensional data, a stepwise least squares algorithm is proposed. This algorithm computes model coefficients and evaluates criteria for the selection of the model order stepwise for AR models of successively decreasing order. Numerical simulations indicate that, with the least squares algorithm, the AR model coefficients and the eigenmodes derived from the coefficients are estimated reliably and that the approximate 95% confidence intervals for the coefficients and eigenmodes are rough approximations of the confidence intervals inferred from the simulations.

Categories and Subject Descriptors: G.3 [**Mathematics of Computing**]: Probability and Statistics—*Markov processes; Multivariate statistics; Statistical computing; Stochastic processes; Time series analysis*; I.6.4 [**Simulation and Modeling**]: Model Validation and Analysis; J.2 [**Computer Applications**]: Physical Sciences and Engineering—*Earth and atmospheric sciences; Mathematics and statistics*

A draft of this paper was written in 1995, while Tapio Schneider was with the Department of Physics at the University of Washington in Seattle. Support, during that time, by a Fulbright grant and by the German National Scholarship Foundation (*Studienstiftung des deutschen Volkes*) is gratefully acknowledged. The paper was completed while Tapio Schneider was with the Atmospheric and Oceanic Sciences Program at Princeton University.
Name: Arnold Neumaier
Affiliation: Institut für Mathematik, Universität Wien
Address: Strudlhofgasse 4, A-1090 Wien, Austria, email: neum@cma.univie.ac.at
Name: Tapio Schneider
Affiliation: Courant Institute of Mathematical Sciences, New York University
Address: 251 Mercer Street, New York, NY 10012, U.S.A., email: tapio@cims.nyu.edu

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or direct commercial advantage and that copies show this notice on the first page or initial screen of a display along with the full citation. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, to republish, to post on servers, to redistribute to lists, or to use any component of this work in other works, requires prior specific permission and/or a fee. Permissions may be requested from Publications Dept, ACM Inc., 1515 Broadway, New York, NY 10036 USA, fax +1 (212) 869-0481, or `permissions@acm.org`.

General Terms: Algorithms, Performance, Theory

Additional Key Words and Phrases: Confidence intervals, eigenmodes, least squares, model identification, order selection, parameter estimation, principal oscillation pattern

1. INTRODUCTION

Dynamical characteristics of a complex system can often be inferred from analyses of a stochastic time series model fitted to observations of the system [Tiao and Box 1981]. In the geosciences, for example, oscillations of a complex system are sometimes characterized by what are known as principal oscillation patterns, eigenmodes of a multivariate autoregressive model of first order [AR(1) model] fitted to observations [Hasselmann 1988; von Storch and Zwiers 1999, Chapter 15]. Principal oscillation patterns possess characteristic frequencies (or oscillation periods) and damping times, the frequencies being the natural frequencies of the AR(1) model. By analyzing principal oscillation patterns of an oscillatory system, one can identify components of the system that are associated with characteristic frequencies and damping times. Xu and von Storch [1990], for example, use a principal oscillation pattern analysis to identify the spatial structures of the mean sea level pressure that are associated with the conglomerate of climatic phenomena collectively called El Niño and the Southern Oscillation. In a similar manner, Huang and Shukla [1997] distinguish those spatial structures of the sea surface temperature that oscillate with periods on the order of years from those that oscillate with periods on the order of decades. More examples of such analyses can be found in the bibliographies of these papers.

Since the principal oscillation pattern analysis is an analysis of eigenmodes of an AR(1) model, dynamical characteristics of a system can be inferred from principal oscillation patterns only if an AR(1) model provides an adequate fit to the observations of the system. The applicability of the principal oscillation pattern analysis is therefore restricted. Generalizing the analysis of eigenmodes of AR(1) models to autoregressive models of arbitrary order p [AR(p) models], we will render the analysis of eigenmodes of AR models applicable to a larger class of systems.

An m-variate AR(p) model for a stationary time series of state vectors $v_\nu \in \mathcal{R}^m$, observed at equally spaced instants ν, is defined by

$$v_\nu = w + \sum_{l=1}^{p} A_l v_{\nu-l} + \varepsilon_\nu, \qquad \varepsilon_\nu = \text{noise}(C), \tag{1}$$

where the m-dimensional vectors $\varepsilon_\nu = \text{noise}(C)$ are uncorrelated random vectors with mean zero and covariance matrix $C \in \mathcal{R}^{m\times m}$, and the matrices $A_1, \ldots, A_p \in \mathcal{R}^{m\times m}$ are the coefficient matrices of the AR model. The parameter vector $w \in \mathcal{R}^m$ is a vector of intercept terms that is included to allow for a nonzero mean of the time series,

$$\langle v_\nu \rangle = (I - A_1 - \cdots - A_p)^{-1} w, \tag{2}$$

where $\langle\cdot\rangle$ denotes an expected value. (For an introduction to modeling multivariate time series with such AR models, see Lütkepohl [1993].) In this paper, we will describe the eigendecomposition of AR(p) models of arbitrary order p. Since the

analysis of eigenmodes of AR models is of interest in particular for high-dimensional systems such as the ones examined in the geosciences, we will also discuss how the order p of an AR model and the coefficient matrices $A_1, \ldots, A_p$, the intercept vector w, and the noise covariance matrix C can be estimated from high-dimensional time series data in a computationally efficient and stable way.

In Section 2, it is shown that an m-variate AR(p) model has mp eigenmodes that possess, just like the m eigenmodes of an AR(1) model, characteristic frequencies and damping times. The excitation is introduced as a measure of the dynamical importance of the modes. Section 3 describes a stepwise least squares algorithm for the estimation of parameters of AR models. This algorithm uses a QR factorization of a data matrix to evaluate, for a sequence of successive orders, a criterion for the selection of the model oder and to compute the parameters of the AR model of the optimum order. Section 4 discusses the construction of approximate confidence intervals for the intercept vector, for the AR coefficients, for the eigenmodes derived from the AR coefficients, and for the oscillation periods and damping times of the eigenmodes. Section 5 contains results of numerical experiments with the presented algorithms. Section 6 summarizes the conclusions.

The methods presented in this paper are implemented in the Matlab package ARFIT, which is described in a companion paper [Schneider and Neumaier 2000]. We will refer to modules in ARFIT that contain implementations of the algorithms under consideration.

Notation. $A_{:k}$ denotes the kth column of the matrix A. A^T is the transpose, and $A^\dagger$ the conjugate transpose of A. The inverse of $A^\dagger$ is written as $A^{-\dagger}$, and the superscript * denotes complex conjugation. In notation, we do not distinguish between random variables and their realizations; whether a symbol refers to a random variable or to a realization can be inferred from the context.

2. EIGENDECOMPOSITION OF AR MODELS

The eigendecomposition of an AR(p) model is a structural analysis of the AR coefficient matrices $A_1, \ldots, A_p$. The eigendecomposition of AR(1) models is described, for example, by Honerkamp [1994, pp. 426 ff.]. In what sense and to what extent an eigendecomposition of an AR(1) model can yield insight into dynamical characteristics of complex systems is discussed by von Storch and Zwiers [1999, Chapter 15]. In Section 2.1, a review of the eigendecomposition of AR(1) models introduces the concepts and notation used throughout this paper. In Section 2.2, the results for AR(1) models are generalized to AR(p) models of arbitrary order p. Throughout this section, we assume that the mean (2) has been subtracted from the time series of state vectors v_ν, so that the intercept vector w can be taken to be zero.

2.1 AR models of first order

Suppose the coefficient matrix A of the m-variate AR(1) model

$$v_\nu = Av_{\nu-1} + \varepsilon_\nu, \qquad \varepsilon_\nu = \text{noise}(C), \tag{3}$$

is nondefective, so that it has m (generally complex) eigenvectors that form a basis of the vector space $\mathcal{R}^m$ of the state vectors v_ν. Let S be the nonsingular matrix that contains these eigenvectors as columns $S_{:k}$, and let $\Lambda = \text{Diag}(\lambda_k)$ be the associated

diagonal matrix of eigenvalues λ_k ($k = 1, \ldots, m$). The eigendecomposition of the coefficient matrix can then be written as $A = S\Lambda S^{-1}$. In the basis of the eigenvectors $S_{:k}$ of the coefficient matrix A, the state vectors v_ν and the noise vectors ε_ν can be represented as linear combinations

$$v_\nu = \sum_{k=1}^{m} v_\nu^{(k)} S_{:k} = S v'_\nu \quad \text{and} \quad \varepsilon_\nu = \sum_{k=1}^{m} \varepsilon_\nu^{(k)} S_{:k} = S \varepsilon'_\nu, \tag{4}$$

with coefficient vectors $v'_\nu = (v_\nu^{(1)}, \ldots, v_\nu^{(m)})^T$ and $\varepsilon'_\nu = (\varepsilon_\nu^{(1)}, \ldots, \varepsilon_\nu^{(m)})^T$. Substituting these expansions of the state vectors v_ν and of the noise vectors ε_ν into the AR(1) model (3) yields, for the coefficient vectors v'_ν, an AR(1) model

$$v'_\nu = \Lambda v'_{\nu-1} + \varepsilon'_\nu, \qquad \varepsilon'_\nu = \text{noise}(C'), \tag{5}$$

with a diagonal coefficient matrix Λ and with a transformed noise covariance matrix

$$C' = S^{-1} C S^{-\dagger}. \tag{6}$$

The m-variate AR(1) model for the coefficient vectors represents a system of m univariate models

$$v_\nu^{(k)} = \lambda_k v_{\nu-1}^{(k)} + \varepsilon_\nu^{(k)}, \qquad k = 1, \ldots, m, \tag{7}$$

which are coupled only via the covariances $\langle \varepsilon_\mu^{(k)} \varepsilon_\nu^{(l)^*} \rangle = \delta_{\mu\nu} C'_{kl}$ of the noise coefficients (where $\delta_{\mu\nu} = 1$ for $\mu = \nu$ and $\delta_{\mu\nu} = 0$ otherwise).

Since the noise vectors are assumed to have mean zero, the dynamics of the expected values of the coefficients

$$\langle v_\nu^{(k)} \rangle = \lambda_k \langle v_{\nu-1}^{(k)} \rangle$$

decouple completely. In the complex plane, the expected values of the coefficients describe a spiral

$$\langle v_{\nu+t}^{(k)} \rangle = \lambda_k^t \langle v_\nu^{(k)} \rangle = e^{-t/\tau_k} e^{(\arg \lambda_k) i t} \langle v_\nu^{(k)} \rangle$$

with *damping time*

$$\tau_k \equiv -1/\log|\lambda_k| \tag{8}$$

and *period*

$$T_k \equiv \frac{2\pi}{|\arg \lambda_k|}, \tag{9}$$

the damping time and the period being measured in units of the sampling interval of the time series v_ν. To render the argument $\arg z = \text{Im}(\log z)$ of a complex number z unique, we stipulate $-\pi \leq \arg z \leq \pi$, a convention that ensures that a pair of complex conjugate eigenvalues is associated with a single period. For a stable AR model with nonsingular coefficient matrix A, the absolute values of all eigenvalues λ_k lie between zero and one, $0 < |\lambda_k| < 1$, which implies that all damping times τ_k of such a model are positive and bounded.

Whether an eigenvalue is real and, if it is real, whether it is positive or negative determines the dynamical character of the eigenmode to which the eigenvalue belongs. If an eigenvalue λ_k has a nonzero imaginary part or if it is real and negative,

the period T_k is bounded, and the AR(1) model (7) for the associated time series of coefficients $v_\nu^{(k)}$ is called a stochastically forced *oscillator*. The period of an oscillator attains its minimum value $T_k = 2$ if the eigenvalue λ_k is real and negative, that is, if the absolute value $|\arg \lambda_k|$ of the argument of the eigenvalue is equal to π. The smallest oscillation period $T_k = 2$ that is representable in a time series sampled at a given sampling interval corresponds to what is known in Fourier analysis as the Nyquist frequency. If an eigenvalue λ_k is real and positive, the period T_k is infinite, and the AR(1) model (7) for the associated time series of coefficients $v_\nu^{(k)}$ is called a stochastically forced *relaxator*.

Thus, a coefficient $v_\nu^{(k)}$ in the expansion (4) of the state vectors v_ν in terms of eigenmodes $S_{:k}$ represents, depending on the eigenvalue λ_k, either a stochastically forced oscillator or a stochastically forced relaxator. The oscillators and relaxators are coupled only via the covariances of the noise, the stochastic forcing. The coefficients $v_\nu^{(k)}$ can be viewed as the amplitudes of the eigenmodes $S_{:k}$ if the eigenmodes are normalized such that $\|S_{:k}\|_2 = 1$. To obtain a unique representation of the eigenmodes $S_{:k}$, we stipulate that the real parts $X = \mathrm{Re}\, S$ and the imaginary parts $Y = \mathrm{Im}\, S$ of the eigenmodes $S_{:k} = X_{:k} + iY_{:k}$ satisfy the normalization conditions

$$X_{:k}^T X_{:k} + Y_{:k}^T Y_{:k} = 1, \qquad X_{:k}^T Y_{:k} = 0, \qquad Y_{:k}^T Y_{:k} < X_{:k}^T X_{:k}. \tag{10}$$

The normalized eigenmodes $S_{:k}$ represent aspects of the system under consideration whose amplitudes $v_\nu^{(k)}$ oscillate with a characteristic period T_k and would, in the absence of stochastic forcing, decay towards zero with a characteristic damping time τ_k. Only oscillatory modes with a finite period have imaginary parts. The real parts and the imaginary parts of oscillatory modes represent aspects of the system under consideration in different phases of an oscillation, with a phase lag of $\pi/2$ between real part and imaginary part. In the geosciences, for example, the state vectors v_ν might represent the Earth's surface temperature field on a spatial grid, with each state vector component representing the temperature at a grid point. The eigenmodes would then represent structures of the surface temperature field that oscillate and decay with characteristic periods and damping times. In a principal oscillation pattern analysis, the spatial structures of the real parts and of the imaginary parts of the eigenmodes are analyzed graphically. It is in this way that, in the geosciences, eigenmodes of AR(1) models are analyzed to infer dynamical characteristics of a complex system (see von Storch and Zwiers [1999, Chapter 15] for more details, including the relationship between the periods of the eigenmodes and the maxima of the power spectrum of an AR model).

Dynamical importance of modes. The magnitudes of the amplitudes $v_\nu^{(k)}$ of the normalized eigenmodes $S_{:k}$ indicate the dynamical importance of the various relaxation and oscillation modes. The variance of an amplitude $v_\nu^{(k)}$, or the *excitation*

$$\sigma_k \equiv \langle |v_\nu^{(k)}|^2 \rangle, \tag{11}$$

is a measure of the dynamical importance of an eigenmode $S_{:k}$.

The excitations can be computed from the coefficient matrix A and from the noise covariance matrix C. The covariance matrix $\Sigma = \langle v_\nu v_\nu^T \rangle$ of the state vectors

v_ν satisfies [Lütkepohl 1993, Chapter 2.1.4]

$$\Sigma = A\Sigma A^T + C. \tag{12}$$

Upon substitution of the eigendecomposition $A = S\Lambda S^{-1}$ and of the transformed noise covariance matrix $C = SC'S^\dagger$, the state covariance matrix becomes

$$\Sigma = S\Sigma' S^\dagger \tag{13}$$

with a transformed state covariance matrix Σ' that is a solution of the linear matrix equation $\Sigma' = \Lambda\Sigma'\Lambda^\dagger + C'$. Since the eigenvalue matrix Λ is diagonal, this matrix equation can be written componentwise as

$$(1 - \lambda_k\lambda_l^*)\Sigma'_{kl} = C'_{kl}.$$

For a stable AR model for which the absolute values of all eigenvalues are less than one, $|\lambda_k| < 1$, this equation can be solved for the transformed state covariance matrix Σ', whose diagonal elements Σ'_{kk} are the excitations σ_k, the variances of the amplitudes $v_\nu^{(k)}$. In terms of the transformed noise covariance matrix C' and of the eigenvalues λ_k, the excitations can hence be written as

$$\sigma_k = \frac{C'_{kk}}{1 - |\lambda_k|^2}, \tag{14}$$

an expression that can be interpreted as the ratio of the forcing strength C'_{kk} over the damping $1 - |\lambda_k|^2$ of an eigenmode $S_{:k}$.

The suggestion of measuring the dynamical importance of the modes in terms of the excitations σ_k contrasts with traditional studies in which the least damped eigenmodes of an AR(1) model were considered dynamically the most important. That is, the eigenmodes for which the associated eigenvalue had the greatest absolute value $|\lambda_k|$ were considered dynamically the most important (e.g., von Storch et al. [1995], Penland and Sardeshmukh [1995], von Storch and Zwiers [1999, Chapter 15], and references therein). The tradition of viewing the least damped modes as the dynamically most important ones comes from the analysis of modes of deterministic linear systems, in which the least damped mode, if excited, dominates the dynamics in the limit of long times. In the presence of stochastic forcing, however, the weakly damped modes, if they are not sufficiently excited by the noise, need not dominate the dynamics in the limit of long times. The excitation σ_k therefore appears to be a more appropriate measure of dynamical importance.

2.2 AR models of arbitrary order

To generalize the eigendecomposition of AR models of first order to AR models of arbitrary order, we exploit the fact that an m-variate AR(p) model

$$v_\nu = \sum_{l=1}^{p} A_l v_{\nu-l} + \varepsilon_\nu, \qquad \varepsilon_\nu = \text{noise}(C),$$

is equivalent to an AR(1) model

$$\tilde{v}_\nu = \tilde{A}\tilde{v}_{\nu-1} + \tilde{\varepsilon}_\nu, \qquad \tilde{\varepsilon} = \text{noise}(\tilde{C}),$$

with augmented state vectors and noise vectors

$$\tilde{v}_\nu = \begin{pmatrix} v_\nu \\ v_{\nu-1} \\ \vdots \\ v_{\nu-p+1} \end{pmatrix} \in \mathcal{R}^{mp} \quad \text{and} \quad \tilde{\varepsilon}_\nu = \begin{pmatrix} \varepsilon_\nu \\ 0 \\ \vdots \\ 0 \end{pmatrix} \in \mathcal{R}^{mp}$$

and with a coefficient matrix [e.g., Honerkamp 1994, p. 426]

$$\tilde{A} = \begin{pmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I & 0 & \cdots & 0 & 0 \\ 0 & I & \cdots & 0 & 0 \\ 0 & 0 & \ddots & 0 & 0 \\ 0 & 0 & \cdots & I & 0 \end{pmatrix} \in \mathcal{R}^{mp \times mp}. \tag{15}$$

The noise covariance matrix

$$\tilde{C} = \langle \tilde{\varepsilon}_\nu \tilde{\varepsilon}_\nu^\dagger \rangle = \begin{pmatrix} C & 0 \\ 0 & 0 \end{pmatrix} \in \mathcal{R}^{mp \times mp}$$

of the equivalent AR(1) model is singular.

An AR(1) model that is equivalent to an AR(p) model can be decomposed into eigenmodes according to the above decomposition of a general AR(1) model. If the augmented coefficient matrix $\tilde{A}$ is nondefective, its mp eigenvectors form a basis of the vector space $\mathcal{R}^{mp}$ of the augmented state vectors $\tilde{v}_\nu$. As above, let $\tilde{S}$ be the nonsingular matrix whose columns $\tilde{S}_{:k}$ are the eigenvectors of the augmented coefficient matrix $\tilde{A} = \tilde{S}\Lambda\tilde{S}^{-1}$, and let Λ be the associated diagonal matrix of eigenvalues λ_k ($k = 1, \ldots, mp$). In terms of the eigenvectors $\tilde{S}_{:k}$ of the augmented coefficient matrix $\tilde{A}$, the augmented state vectors and noise vectors can be represented as linear combinations

$$\tilde{v}_\nu = \sum_{k=1}^{mp} \tilde{v}_\nu^{(k)} \tilde{S}_{:k} \quad \text{and} \quad \tilde{\varepsilon}_\nu = \sum_{k=1}^{mp} \tilde{\varepsilon}_\nu^{(k)} \tilde{S}_{:k}. \tag{16}$$

The dynamics of the coefficients $\tilde{v}_\nu^{(k)}$ in the expansion of the augmented state vectors are governed by a system of mp univariate AR(1) models

$$\tilde{v}_\nu^{(k)} = \lambda_k \tilde{v}_{\nu-1}^{(k)} + \tilde{\varepsilon}_\nu^{(k)}, \qquad k = 1, \ldots, mp, \tag{17}$$

which are coupled only via the covariances $\langle \tilde{\varepsilon}_\nu^{(k)} \tilde{\varepsilon}_\mu^{(l)^*} \rangle = \delta_{\mu\nu} \tilde{C}'_{kl}$ of the noise coefficients. The covariance matrix of the noise coefficients is the transformed noise covariance matrix

$$\tilde{C}' = \tilde{S}^{-1} \tilde{C} \tilde{S}^{-\dagger} \tag{18}$$

of the equivalent AR(1) model. Thus, the augmented time series $\tilde{v}_\nu$ can be decomposed, just as above, into mp oscillators and relaxators with mp-dimensional eigenmodes $\tilde{S}_{:k}$.

However, because the augmented coefficient matrix $\tilde{A}$ has the Frobenius structure (15), the augmented eigenmodes $\tilde{S}_{:k}$ have a structure that makes it possible

to decompose the original time series v_ν into oscillators and relaxators with m-dimensional modes, instead of the augmented mp-dimensional modes $\tilde{S}_{:k}$. The eigenvectors $\tilde{S}_{:k}$ of the augmented coefficient matrix $\tilde{A}$ have the structure

$$\tilde{S}_{:k} = \begin{pmatrix} \lambda_k^{p-1} S_{:k} \\ \vdots \\ \lambda_k S_{:k} \\ S_{:k} \end{pmatrix}$$

with an m-dimensional vector $S_{:k}$ (cf. Wilkinson's [1965, Chapter 1.30] discussion of eigenmodes of higher-order differential equations). Substituting this expression for the augmented eigenvectors $\tilde{S}_{:k}$ into the expansions (16) of the augmented state vectors and noise vectors and introducing the renormalized coefficients

$$v_\nu^{(k)} \equiv \lambda_k^{p-1} \tilde{v}_\nu^{(k)} \quad \text{and} \quad \varepsilon_\nu^{(k)} \equiv \lambda_k^{p-1} \tilde{\varepsilon}_\nu^{(k)},$$

one finds that the original m-dimensional state vectors v_ν and noise vectors ε_ν can be represented as linear combinations

$$v_\nu = \sum_{k=1}^{mp} v_\nu^{(k)} S_{:k} \quad \text{and} \quad \varepsilon_\nu = \sum_{k=1}^{mp} \varepsilon_\nu^{(k)} S_{:k} \tag{19}$$

of the m-dimensional vectors $S_{:k}$. Like the dynamics (17) of the coefficients $\tilde{v}_\nu^{(k)}$ in the expansion of the augmented state vectors $\tilde{v}_\nu$, the dynamics of the coefficients $v_\nu^{(k)}$ in the expansion of the original state vectors v_ν are governed by a system of mp univariate AR(1) models

$$v_\nu^{(k)} = \lambda_k v_{\nu-1}^{(k)} + \varepsilon_\nu^{(k)}, \qquad k = 1, \ldots, mp, \tag{20}$$

which are coupled only via the covariances $\langle \varepsilon_\mu^{(k)} \varepsilon_\nu^{(l)*} \rangle = \delta_{\mu\nu} (\lambda_k \lambda_l^*)^{p-1} \tilde{C}'_{kl}$ of the noise coefficients.

For AR(p) models of arbitrary order, the expansions (19) of the state vectors v_ν and of the noise vectors ε_ν and the dynamics (20) of the expansion coefficients parallel the expansions (4) of the state vectors and of the noise vectors and the dynamics (7) of the expansion coefficients for AR(1) models. In the decomposition of an AR(p) model of arbitrary order, the AR(1) models (20) for the dynamics of the expansion coefficients $v_\nu^{(k)}$ can be viewed, as in the decomposition of AR(1) models, as oscillators or relaxators, depending on the eigenvalue λ_k of the augmented coefficient matrix $\tilde{A}$. The m-dimensional vectors $S_{:k}$ can be viewed as eigenmodes that possess characteristic damping times (8) and periods (9). To obtain a unique representation of the eigenmodes $S_{:k}$, we stipulate, as an extension of the normalization (10) in the first-order case, that the real parts $\tilde{X} = \operatorname{Re} \tilde{S}$ and the imaginary parts $\tilde{Y} = \operatorname{Im} \tilde{S}$ of the eigenvectors $\tilde{S}_{:k} = \tilde{X}_{:k} + i\tilde{Y}_{:k}$ of the augmented coefficient matrix $\tilde{A}$ satisfy the normalization conditions

$$\tilde{X}_{:k}^T \tilde{X}_{:k} + \tilde{Y}_{:k}^T \tilde{Y}_{:k} = 1, \qquad \tilde{X}_{:k}^T \tilde{Y}_{:k} = 0, \qquad \tilde{Y}_{:k}^T \tilde{Y}_{:k} < \tilde{X}_{:k}^T \tilde{X}_{:k}. \tag{21}$$

With this normalization of the eigenvectors $\tilde{S}_{:k}$, the coefficients $\tilde{v}_\nu^{(k)}$ in the expansion of the augmented state vectors $\tilde{v}_\nu$ indicate the amplitudes of the modes $S_{:k}$. The variances of these amplitudes, the excitations $\sigma_k = \langle |\tilde{v}_\nu^{(k)}|^2 \rangle = |\lambda_k|^{2(1-p)} \langle |v_\nu^{(k)}|^2 \rangle$,

are measures of the dynamical importance of the modes. In analogy to the excitations (14) of modes of AR(1) models, the excitations of modes of AR(p) models can be expressed as

$$\sigma_k = \frac{\tilde{C}'_{kk}}{1 - |\lambda_k|^2},$$

where $\tilde{C}'_{kk}$ is a diagonal element of the transformed noise covariance matrix (18).

Thus, an AR(p) model of arbitrary order can be decomposed, just like an AR(1) model, into mp oscillators and relaxators with m-dimensional eigenmodes $S_{:k}$. To infer dynamical characteristics of a complex system, the eigenmodes of AR(p) models can be analyzed in the same way as the eigenmodes of AR(1) models. All results for AR(1) models have a direct extension to AR(p) models. The only difference between the eigendecomposition of AR(1) models and the eigendecomposition of higher-order AR(p) models is that higher-order models possess a larger number of eigenmodes, which span the vector space of the state vectors v_ν but are not, as in the first-order case, linearly independent.

The ARFIT module **armode** computes the eigenmodes $S_{:k}$ of AR(p) models of arbitrary order by an eigendecomposition of the coefficient matrix $\tilde{A}$ of an equivalent AR(1) model. It also computes the periods, damping times, and excitations of the eigenmodes.

3. STEPWISE LEAST SQUARES ESTIMATION FOR AR MODELS

To analyze the eigenmodes of an AR(p) model fitted to a time series of observations of a complex system, the unknown model order p and the unknown model parameters $A_1, \ldots, A_p$, w, and C must first be estimated. The model order is commonly estimated as the optimizer of what is called an order selection criterion, a function that depends on the noise covariance matrix $\hat{C}$ of an estimated AR(p) model and that penalizes the overparameterization of a model [Lütkepohl 1993, Chapter 4.3; the hat-accent $\hat{A}$ designates an estimate of the quantity A]. To determine the model order p_{opt} that optimizes the order selection criterion, the noise covariance matrices C are estimated and the order selection criterion is evaluated for AR(p) models of successive orders $p_{\min} \le p \le p_{\max}$. If the parameters $A_1, \ldots, A_p$, and w are not estimated along with the noise covariance matrix C, they are then estimated for a model of the optimum order p_{opt}.

Both asymptotic theory and simulations indicate that, if the coefficient matrices $A_1, \ldots, A_p$ and the intercept vector w of an AR model are estimated with the method of least squares, the residual covariance matrix $\hat{C}$ of the estimated model is a fairly reliable estimator of the noise covariance matrix C and hence can be used in order selection criteria [Tjøstheim and Paulsen 1983; Paulsen and Tjøstheim 1985; Mentz et al. 1998].[1] The least squares estimates of AR parameters are obtained by casting an AR model in the form of an ordinary regression model and estimating the parameters of the regression model with the method of least squares [Lütkepohl 1993, Chapter 3]. Numerically, the least squares problem for the ordinary regression

[1] Although the residual covariance matrix of an AR model whose parameters are estimated with the method of least squares is not itself a least squares estimate of the noise covariance matrix, we will, as is common practice, refer to this residual covariance matrix as a least squares estimate.

model can be solved with standard methods that involve the factorization of a data matrix [e.g., Björck 1996, Chapter 2]. In what follows, we will present a stepwise least squares algorithm with which, in a computationally efficient and stable manner, the parameters of an AR model can be estimated and an order selection criterion can be evaluated for AR(p) models of successive orders $p_{\min} \leq p \leq p_{\max}$. Starting from a review of how the least squares estimates for an AR(p) model of fixed order p can be computed via a QR factorization of a data matrix, we will show how, from the same QR factorization, approximate least squares estimates for models of lower order $p' < p$ can be obtained.

3.1 Least squares estimates for an AR model of fixed order

Suppose an m-dimensional time series of $N+p$ state vectors v_ν ($\nu = 1-p, \ldots, N$) is available, the time series consisting of p pre-sample state vectors $v_{1-p}, \ldots, v_0$ and N state vectors $v_1, \ldots, v_N$ that form what we call the effective sample. The parameters $A_1, \ldots, A_p$, w, and C of an AR(p) model of fixed order p are to be estimated.

An AR(p) model can be cast in the form of a regression model

$$v_\nu = Bu_\nu + \varepsilon_\nu, \qquad \varepsilon_\nu = \text{noise}(C), \qquad \nu = 1, \ldots, N, \tag{22}$$

with parameter matrix

$$B = (w \; A_1 \; \cdots \; A_p) \tag{23}$$

and with predictors

$$u_\nu = \begin{pmatrix} 1 \\ v_{\nu-1} \\ \vdots \\ v_{\nu-p} \end{pmatrix} \tag{24}$$

of dimension $n_p = mp+1$. Casting an AR model in the form of a regression model is an approximation in that in a regression model, the predictors u_ν are assumed to be constant, whereas the state vectors v_ν of an AR process are a realization of a stochastic process. The approximation of casting an AR model into the form of a regression model amounts to treating the first predictor

$$u_1 = \begin{pmatrix} 1 \\ v_0 \\ \vdots \\ v_{1-p} \end{pmatrix}$$

as a vector of constant initial values [cf. Wei 1994, Chapter 7.2.1]. What are unconditional parameter estimates for the regression model are therefore conditional parameter estimates for the AR model, conditional on the first p pre-sample state vectors $v_{1-p}, \ldots, v_0$ being constant. But since the relative influence of the initial condition on the parameter estimates decreases as the sample size N increases, the parameter estimates for the regression model are still consistent and asymptotically unbiased estimates for the AR model [e.g., Lütkepohl 1993, Chapter 3].

In terms of the moment matrices

$$U = \sum_{\nu=1}^{N} u_\nu u_\nu^T, \qquad V = \sum_{\nu=1}^{N} v_\nu v_\nu^T, \qquad W = \sum_{\nu=1}^{N} v_\nu u_\nu^T, \tag{25}$$

the least squares estimate of the parameter matrix B can be written as

$$\hat{B} = WU^{-1}. \tag{26}$$

The residual covariance matrix

$$\hat{C} = \frac{1}{N - n_p} \sum_{\nu=1}^{N} \hat{\varepsilon}_\nu \hat{\varepsilon}_\nu^T \quad \text{with} \quad \hat{\varepsilon}_\nu = v_\nu - \hat{B} u_\nu$$

is an estimate of the noise covariance matrix C and can be expressed as

$$\hat{C} = \frac{1}{N - n_p}(V - WU^{-1}W^T). \tag{27}$$

A derivation of the least squares estimators and a discussion of their properties can be found, for example, in Lütkepohl [1993, Chapter 3].

The residual covariance matrix $\hat{C}$ is proportional to a Schur complement of the matrix

$$\Gamma = \begin{pmatrix} U & W^T \\ W & V \end{pmatrix} = \sum_{\nu=1}^{N} \begin{pmatrix} u_\nu \\ v_\nu \end{pmatrix} \begin{pmatrix} u_\nu^T & v_\nu^T \end{pmatrix},$$

which is the moment matrix $\Gamma = K^T K$ belonging to the data matrix

$$K = \begin{pmatrix} u_1^T & v_1^T \\ \vdots & \vdots \\ u_N^T & v_N^T \end{pmatrix}. \tag{28}$$

The least squares estimates can be computed from a QR factorization of the data matrix

$$K = QR, \tag{29}$$

with an orthogonal matrix Q and an upper triangular matrix

$$R = \begin{pmatrix} R_{11} & R_{12} \\ 0 & R_{22} \end{pmatrix}.$$

The QR factorization of the data matrix K leads to the Cholesky factorization $\Gamma = K^T K = R^T R$ of the moment matrix,

$$\begin{pmatrix} U & W^T \\ W & V \end{pmatrix} = R^T R = \begin{pmatrix} R_{11}^T R_{11} & R_{11}^T R_{12} \\ R_{12}^T R_{11} & R_{12}^T R_{12} + R_{22}^T R_{22} \end{pmatrix}, \tag{30}$$

and from this Cholesky factorization, one finds the representation

$$\hat{B} = \left(R_{11}^{-1} R_{12}\right)^T \quad \text{and} \quad \hat{C} = \frac{1}{N - n_p} R_{22}^T R_{22} \tag{31}$$

for the least squares estimates of the parameter matrix B and of the noise covariance matrix C. The estimated parameter matrix $\hat{B}$ is obtained as the solution of a

triangular system of equations, and the residual covariance matrix $\hat{C}$ is given in a factored form that shows explicitly that the residual covariance matrix is positive semidefinite.

If the moment matrix $\Gamma = K^T K$ is ill-conditioned, the effect of rounding errors and data errors on the parameter estimates can be reduced by computing the parameter estimates (31) not from a Cholesky factorization (30) of the moment matrix Γ, but from an analogous Cholesky factorization of a regularized moment matrix $\Gamma + \delta D^2$, where D^2 is a positive definite diagonal matrix and δ is a regularization parameter [e.g., Hansen 1997]. A Cholesky factorization of the regularized moment matrix $\Gamma + \delta D^2 = R^T R$ can be computed via a QR factorization of the augmented data matrix

$$\begin{pmatrix} K \\ \sqrt{\delta} D \end{pmatrix} = QR. \tag{32}$$

Rounding errors and data errors have a lesser effect on the estimates (31) computed from the upper triangular factor R of this QR factorization. The diagonal matrix D might be chosen to consist of the Euclidean norms of the columns of the data matrix, $D = \mathrm{Diag}\big(\|K_{:j}\|_2\big)$. The regularization parameter δ, as a heuristic, might be chosen to be a multiple $(q^2+q+1)\eta$ of the machine precision η, the multiplication factor q^2+q+1 depending on the dimension $q = n_p+m$ of the moment matrix Γ (cf. Higham's [1996] Theorem 10.7, which implies that with such a regularization the direct computation of a Cholesky factorization of the regularized moment matrix $\Gamma+\delta D^2$ would be well-posed). The ARFIT module **arqr** computes such a regularized QR factorization for AR models.

If the observational error of the data is unknown but dominates the rounding error, the regularization parameter can be estimated with adaptive regularization techniques. In this case, however, the QR factorization (32) should be replaced by a singular value decomposition of the rescaled data matrix KD^{-1}, because the singular value decomposition can be used more efficiently with adaptive regularization methods [see, e.g., Hansen 1997; Neumaier 1998].

3.2 Downdating the least squares estimates

To select the order of an AR model, the residual covariance matrix $\hat{C}$ must be computed and an order selection criterion must be evaluated for AR(p) models of successive orders $p_{\min} \le p \le p_{\max}$. Order selection criteria are usually functions of the logarithm of the determinant

$$l_p = \log\det \Delta_p$$

of the residual cross-product matrix

$$\Delta_p = (N - n_p)\hat{C} = R_{22}^T R_{22}.$$

For example, Schwarz's [1978] Bayesian Criterion (SBC) can be written as

$$\mathrm{SBC}(p) = \frac{l_p}{m} - \Big(1 - \frac{n_p}{N}\Big)\log N,$$

and the logarithm of Akaike's [1971] Final Prediction Error (FPE) criterion as

$$\mathrm{FPE}(p) = \frac{l_p}{m} - \log\frac{N(N-n_p)}{N+n_p}.$$

(These and other order selection criteria and their properties are discussed, for example, by Lütkepohl [1985; 1993, Chapter 4].) Instead of computing the residual cross-product matrix Δ_p by a separate QR factorization for each order p for which an order selection criterion is to be evaluated, one can compute an approximation of the residual cross-product matrix Δ_p for a model of order $p < p_{\max}$ by downdating the QR factorization for a model of order $p_{\max}$. (For a general discussion of updating and downdating least squares problems, see Björck [1996, Chapter 3].)

To downdate the QR factorization for a model of order p to a structurally similar factorization for a model of order $p' = p - 1$, one exploits the structure

$$K = (K_1\ K_2), \qquad K_1 = \begin{pmatrix} u_1^T \\ \vdots \\ u_N^T \end{pmatrix}, \qquad K_2 = \begin{pmatrix} v_1^T \\ \vdots \\ v_N^T \end{pmatrix}$$

of the data matrix (28). A data matrix $K' = (K_1'\ K_2)$ for a model of order $p' = p-1$ follows, approximately, from the data matrix $K = (K_1\ K_2)$ for a model of order p by removing from the submatrix $K_1 = (K_1'\ K_1'')$ of the predictors u_ν the m trailing columns K_1''. The downdated data matrix K' is only approximately equal to the least squares data matrix (28) for a model of order p' because in the downdated data matrix K', the first available state vector v_{1-p} is not included. When parameter estimates are computed from the downdated data matrix K', a sample of the same effective size N is assumed both for the least squares estimates of order p and the least squares estimates of order $p' = p - 1$, although in the case of the lower order p', the pre-sample state vector v_0 could become part of the effective sample so that a sample of effective size $N + 1$ would be available. The relative loss of accuracy that this approximation entails decreases with increasing sample size N.

A factorization of the downdated data matrix K' follows from the QR factorization of the original data matrix K if one partitions the submatrices R_{11} and R_{12} of the triangular factor R, considering the m trailing rows and columns of R_{11} and the m trailing rows of R_{12} separately,

$$R_{11} = \begin{pmatrix} R_{11}' & R_{11}'' \\ 0 & R_{11}''' \end{pmatrix}, \qquad R_{12} = \begin{pmatrix} R_{12}' \\ R_{12}'' \end{pmatrix}.$$

With the thus partitioned triangular factor R, the QR factorization (29) of the data matrix becomes

$$K = (K_1'\ K_1''\ K_2) = Q \begin{pmatrix} R_{11}' & R_{11}'' & R_{12}' \\ 0 & R_{11}''' & R_{12}'' \\ 0 & 0 & R_{22} \end{pmatrix}.$$

Dropping the m columns belonging to the submatrix K_1'', one obtains a factorization of the downdated data matrix

$$K' = (K_1'\ K_2) = Q \begin{pmatrix} R_{11}' & R_{12}' \\ 0 & R_{12}'' \\ 0 & R_{22} \end{pmatrix} = Q \begin{pmatrix} R_{11}' & R_{12}' \\ 0 & R_{22}' \end{pmatrix}$$

where

$$R_{22}' = \begin{pmatrix} R_{12}'' \\ R_{22} \end{pmatrix}.$$

This downdated factorization has the same block-structure as the QR factorization (29) of the original data matrix K, but the submatrix R'_{22} in the downdated factorization is not triangular. The factorization of the downdated data matrix K' thus is not a QR factorization. That the submatrix R'_{22} is not triangular, however, does not affect the form of the least squares estimates (31), which, for a model of order $p' = p - 1$, can be computed from the downdated factorization in the same way as they are computed from the original QR factorization for a model of order p.

From the downdated factorization of the data matrix, one can obtain, for the evaluation of order selection criteria, downdating formulas for the logarithm $l_p = \log\det\Delta_p$ of the determinant of the residual cross-product matrix Δ_p. The factorization of the downdated data matrix leads to the residual cross-product matrix

$$\Delta_{p'} = R'^T_{22} R'_{22} = R^T_{22} R_{22} + R''^T_{12} R''_{12},$$

from which, with the notation

$$R_p = R''_{12},$$

the downdating formula

$$\Delta_{p-1} = \Delta_p + R_p^T R_p$$

for the residual cross-product matrix follows. Because the determinant of the right-hand side of this formula can be brought into the form [Anderson 1984, Theorem A.3.2]

$$\det(\Delta_p + R_p^T R_p) = \det\Delta_p \cdot \det(I + R_p \Delta_p^{-1} R_p^T),$$

the downdating formula for the determinant term $l_p = \log\det\Delta_p$ becomes

$$l_{p-1} = l_p + \log\det(I + R_p \Delta_p^{-1} R_p^T).$$

This downdate can be computed from a Cholesky factorization

$$I + R_p \Delta_p^{-1} R_p^T = L_p L_p^T \tag{33}$$

as

$$l_{p-1} = l_p + 2\log\det L_p, \tag{34}$$

the determinant of the Cholesky factor L_p being the product of the diagonal elements.

This downdating procedure can be iterated, starting from a QR factorization for a model of order $p_{\max}$ and stepwise downdating the factorization and the determinant term l_p appearing in the order selection criteria until the minimum order $p_{\min}$ is reached. To downdate the inverse cross-product matrix Δ_p^{-1}, which is needed in the Cholesky factorization (33), one can use the Woodbury formula [Björck 1996, Chapter 3]

$$\Delta_{p-1}^{-1} = (\Delta_p + R_p^T R_p)^{-1} = \Delta_p^{-1} - \Delta_p^{-1} R_p^T (I + R_p \Delta_p^{-1} R_p^T)^{-1} R_p \Delta_p^{-1}$$

and compute the downdated inverse Δ_{p-1}^{-1} from the Cholesky factorization (33) via

$$N_p = L_p^{-1} R_p \Delta_p^{-1}, \tag{35}$$

$$\Delta_{p-1}^{-1} = \Delta_p^{-1} - N_p^T N_p. \tag{36}$$

With the downdating scheme (33–36), an order selection criterion such as SBC or FPE, given a QR factorization for a model of order $p_{\max}$, can be evaluated for models of order $p_{\max} - 1, \ldots, p_{\min}$, whereby for the model of order p, the first $p_{\max} - p$ state vectors are ignored.

After evaluating the order selection criterion for a sequence of models and determining an optimum order p_{opt}, one finds the least squares parameter estimates (31) for the model of the optimum order by replacing the maximally sized submatrices R_{11} and R_{12} of the triangular factor R by their leading submatrices of size $n_{p_{\text{opt}}}$. If the available time series is short and the assumed maximum model order $p_{\max}$ is much larger than the selected optimum order p_{opt}, computing the parameters of the AR model of optimum order p_{opt} from the downdated factorization of the model of maximum order $p_{\max}$, and thus ignoring $(p_{\max} - p_{\text{opt}})$ available state vectors, might entail a significant loss of information. To improve the accuracy of the parameter estimates in such cases, the parameters of the model of optimum order p_{opt} can be computed from a second QR factorization, a QR factorization for a model of order $p = p_{\text{opt}}$.

The above downdating scheme is applicable both to the QR factorization of the data matrix K and to the regularized QR factorization of the augmented data matrix (32). The ARFIT module **arord** evaluates order selection criteria by downdating the regularized QR factorization performed with the module **arqr**. The driver module **arfit** determines the optimum model order and computes the AR parameters for the model of the optimum order.

3.3 Computational complexity of the stepwise least squares algorithm

The data matrix whose QR factorization is to be computed is of size $N' \times (n_{p_{\max}} + m)$, where the number of rows N' of this matrix is equal to the sample size N if the least squares estimates are not regularized, or the number of rows N' is equal to $N + n_{p_{\max}} + m$ if the least squares estimates are regularized by computing the QR factorization of the augmented data matrix (32). Computing the QR factorization requires, to leading order, $O(N' m^2 p_{\max}^2)$ operations.

In traditional algorithms for estimating parameters of AR models, a separate factorization would be computed for each order p for which an order selection criterion is to be evaluated. In the stepwise least squares algorithm, the downdates (33)–(36) require $O(m^3)$ operations for each order p for which an order selection criterion is to be evaluated. Since $N' \geq m$, the downdating process for each order p requires fewer operations than a new QR factorization. If the number of rows N' of the data matrix whose QR factorization is computed is much greater than the dimension m of the state vectors, the number of operations required for the downdating process becomes negligible compared with the number of operations required for the QR factorization. With the stepwise least squares algorithm, then, the order and the parameters of an AR model can be estimated about $(p_{\max} - p_{\min} + 1)$-times faster than with traditional least squares algorithms that require $(p_{\max} - p_{\min} + 1)$ separate QR factorizations. Since deleting columns of a matrix does not decrease the smallest singular value of the matrix, the stepwise least squares algorithm is a numerically stable procedure [cf. Björck 1996, Chapter 3.2].

4. CONFIDENCE INTERVALS

Under weak conditions on the distribution of the noise vectors ε_ν of the AR model, the least squares estimator of the AR coefficient matrices $A_1, \ldots, A_p$ and of the intercept vector w is consistent and asymptotically normal [e.g., Lütkepohl 1993, Chapter 3.2]. Let the AR parameters $B = (w\ A_1\ \cdots\ A_p) \in \mathcal{R}^{m \times n_p}$ be arranged into a parameter vector x_B by stacking adjacent columns $B_{:j}$ of the parameter matrix B,

$$x_B = \begin{pmatrix} B_{:1} \\ \vdots \\ B_{:n_p} \end{pmatrix} \in \mathcal{R}^{mn_p}.$$

Asymptotically, in the limit of large sample sizes N, the estimation errors $x_{\hat{B}} - x_B$ are normally distributed with mean zero and with a covariance matrix Σ_B that depends on the noise covariance matrix C and on the moment matrix $\langle u_\nu u_\nu^T \rangle$ of the predictors u_ν in the regression model (22) [e.g., Lütkepohl 1993, Chapter 3.2.2]. Substituting the least squares estimate $\hat{C}$ of the noise covariance matrix and the sample moment matrix U of the predictors u_ν for the unknown population quantities, one obtains, for the least squares estimator $x_{\hat{B}}$, the covariance matrix estimate

$$\hat{\Sigma}_B = U^{-1} \otimes \hat{C}, \tag{37}$$

where $A \otimes B$ denotes the Kronecker product of the matrices A and B.

Basing inferences for finite samples on the asymptotic distribution of the least squares estimator, one can establish approximate confidence intervals for the AR coefficients, for the intercept vector, and for the eigenmodes, periods, and damping times derived from the AR coefficients. A confidence interval for an element $\phi \equiv (x_B)_l$ of the parameter matrix B can be constructed from the distribution of the t-ratio

$$t = \frac{\phi_\pm}{\hat{\sigma}_\phi}, \tag{38}$$

the ratio of the estimation error $\phi_\pm \equiv \hat{\phi} - \phi$ of the least squares estimate $\hat{\phi} = (x_{\hat{B}})_l = \hat{B}_{jk}$ over the square root of the estimated variance

$$\hat{\sigma}_\phi^2 \equiv (\hat{\Sigma}_B)_{ll} = (U^{-1})_{kk} \hat{C}_{jj}, \qquad l = m(k-1) + j, \tag{39}$$

of the least squares estimator. For the construction of confidence intervals, it is common practice to assume that the t-ratio (38) follows Student's t distribution with $N - n_p$ degrees of freedom [e.g., Lütkepohl 1993, Chapter 3.2]. To be sure, this assumption is justified only asymptotically, but for lack of finite-sample statistics, it is commonly made. Assuming a t distribution for the t-ratios yields, for the parameter ϕ, the $100\alpha\%$ confidence limits $\hat{\phi} \pm \hat{\phi}_\pm$ with margin of error

$$\hat{\phi}_\pm = t\big(N - n_p, (1+\alpha)/2\big)\, \hat{\sigma}_\phi, \tag{40}$$

where $t(d, \beta)$ is the β-quantile of a t distribution with d degrees of freedom [cf. Draper and Smith 1981, Chapter 1.4]. From the estimated variance (39) of the least squares estimator, one finds that the margin of error of a parameter estimate

$\hat{\phi} = (x_{\hat{B}})_l = \hat{B}_{jk}$ takes the form

$$\hat{\phi}_\pm = t\big(N - n_p, (1+\alpha)/2\big)\,\sqrt{(U^{-1})_{kk}\hat{C}_{jj}}. \tag{41}$$

The ARFIT module **arconf** uses this expression to compute approximate confidence limits $\hat{\phi} \pm \hat{\phi}_\pm$ for the elements of the AR coefficient matrices and of the intercept vector.

Establishing confidence intervals for the eigenmodes and their periods and damping times is complicated by the fact that these quantities are nonlinear functions of the AR coefficients. Whereas for certain random matrices — for symmetric Wishart matrices, for example [Anderson 1984, Chapter 13] — some properties of the distributions of eigenvectors and eigenvalues are known, no analytical expressions for the distributions of eigenvalues and eigenvectors appear to be known for nonsymmetric Gaussian random matrices with correlated elements. For the estimation of confidence intervals for the eigenmodes and their periods and damping times, we must therefore resort to additional approximations that go beyond the asymptotic approximation invoked in constructing the approximate confidence intervals for the AR coefficients.

Consider a real-valued function $\phi = \phi(x_B)$ that depends continuously on the AR parameters x_B, and let the estimate $\hat{\phi} \equiv \phi(x_{\hat{B}})$ be the value of the function ϕ at the least squares estimates $x_{\hat{B}}$ of the AR parameters x_B. The function ϕ may be, for example, the real part or the imaginary part of a component of an eigenmode, or a period or damping time associated with an eigenmode. Linearizing the function ϕ about the estimate $\hat{\phi}$ leads to $\hat{\phi} - \phi \approx \big(\nabla\hat{\phi}\big)^T(x_{\hat{B}} - x_B)$, where $\nabla\hat{\phi}$ denotes the gradient of $\phi(x_B)$ at the estimate $x_B = x_{\hat{B}}$. From this linearization, it follows that the variance $\sigma_\phi^2 \equiv \big\langle(\hat{\phi} - \phi)^2\big\rangle$ of the estimator function $\hat{\phi}$ can be approximated as

$$\sigma_\phi^2 \approx \big\langle \big(\nabla\hat{\phi}\big)^T(x_{\hat{B}} - x_B)(x_{\hat{B}} - x_B)^T\nabla\hat{\phi}\big\rangle = \big(\nabla\hat{\phi}\big)^T\Sigma_B\big(\nabla\hat{\phi}\big), \tag{42}$$

where Σ_B is the covariance matrix of the least squares estimator $x_{\hat{B}}$. If the function ϕ is linear in the parameters x_B, the relation (42) between the variance σ_ϕ^2 of the estimator function $\hat{\phi}$ and the covariance matrix Σ_B of the estimator $x_{\hat{B}}$ holds exactly. But if the function ϕ is nonlinear, as it is, for example, when ϕ stands for a component of an eigenmode, the relation (42) holds only approximately, up to higher-order terms.

Substituting the asymptotic covariance matrix (37) into the expression (42) for the variance of the estimator function $\hat{\phi}$ gives a variance estimate

$$\hat{\sigma}_\phi^2 = \big(\nabla\hat{\phi}\big)^T\hat{\Sigma}_B\big(\nabla\hat{\phi}\big) \tag{43}$$

that can be used to establish confidence intervals. If the function ϕ is nonlinear, the t-ratio $t = \phi_\pm/\hat{\sigma}_\phi^2$ of the estimation error $\phi_\pm = \hat{\phi} - \phi$ over the square root of the estimated variance $\hat{\sigma}_\phi^2$ generally does not follow a t distribution, not even asymptotically. But assuming that the t-ratio follows a t distribution is still a plausible heuristic for constructing approximate confidence intervals. Generalizing the above construction of confidence limits for the AR parameters, we therefore compute approximate confidence limits $\hat{\phi} \pm \hat{\phi}_\pm$ for functions $\phi(x_B)$ of the AR parameters with the estimator variance (43) and with the margin of error (40).

The ARFIT module **armode** thus establishes approximate confidence intervals for the real parts and the imaginary parts of the individual components of the eigenmodes, and for the periods and the damping times associated with the eigenmodes. The closed-form expressions for the gradients $\nabla\hat{\phi}$ of the eigenmodes, periods, and damping times, which are required for the computation of the estimator variance (43), are derived in the Appendix.

5. NUMERICAL EXAMPLE

To illustrate the least squares estimation of AR parameters and to test the quality of the approximate confidence intervals for the AR parameters and for the eigenmodes, periods, and damping times, we generated time series data by simulation of the bivariate AR(2) process

$$v_\nu = w + A_1 v_{\nu-1} + A_2 v_{\nu-2} + \varepsilon_\nu, \qquad \varepsilon_\nu = \mathrm{WN}(0, C), \qquad \nu = 1, \ldots, N \tag{44}$$

with intercept vector

$$w = \begin{pmatrix} 0.25 \\ 0.10 \end{pmatrix}, \tag{45}$$

coefficient matrices

$$A_1 = \begin{pmatrix} 0.40 & 1.20 \\ 0.30 & 0.70 \end{pmatrix}, \quad A_2 = \begin{pmatrix} 0.35 & -0.30 \\ -0.40 & -0.50 \end{pmatrix}, \tag{46}$$

and noise covariance matrix

$$C = \begin{pmatrix} 1.00 & 0.50 \\ 0.50 & 1.50 \end{pmatrix}. \tag{47}$$

The pseudo-random vectors $\varepsilon_\nu = \mathrm{WN}(0, C)$ are simulated Gaussian white noise vectors with mean zero and covariance matrix C. Ensembles of time series of effective length $N = 25$, 50, 100, and 400 were generated, the ensembles consisting of 20000 time series for $N = 25$; 10000 time series for $N = 50$; and 5000 time series for $N = 100$ and $N = 400$. With the methods of the preceding sections, the AR(2) parameters were estimated from the simulated time series, the eigendecomposition of the estimated models was computed, and approximate 95% confidence intervals were constructed for the AR parameters and for the eigenmodes, periods, and damping times.

Table 1 shows, for each AR parameter $\phi = B_{jk}$, the median of the least squares parameter estimates $\hat{\phi} = \hat{B}_{jk}$ and the median of the margins of error $\hat{\phi}_\pm$ belonging to the approximate 95% confidence intervals (41). Included in the table are the absolute values of the 2.5th percentile ϕ_- and of the 97.5th percentile ϕ_+ of the simulated distribution of the estimation errors $\hat{\phi} - \phi$. 95% of the least squares parameter estimates $\hat{\phi}$ lie between the limits $\phi - \phi_-$ and $\phi + \phi_+$. The quantity

$$q_{95} = 95\text{th percentile of } \left\{ \frac{(\phi_+, \phi_-)}{\hat{\phi}_\pm} \right\}$$

is the 95th percentile of the ratio of the simulated margins of error ϕ_+ and ϕ_- over the approximate margins of error $\hat{\phi}_\pm$. The quantity q_{95} is a measure of how much the approximate margin of error $\hat{\phi}_\pm$ can underestimate the simulated margins of error ϕ_- and ϕ_+.

Table 1. Least squares estimates and 95% margins of error for the parameters of the bivariate AR(2) model (44).

ϕ	$\hat{\phi} \pm \hat{\phi}_\pm$	ϕ_-	ϕ_+	q_{95}	$\hat{\phi} \pm \hat{\phi}_\pm$	ϕ_-	ϕ_+	q_{95}
	$N = 25$				$N = 50$			
w_1	0.289 ± 0.506	0.451	0.820	2.13	0.266 ± 0.328	0.289	0.453	1.67
w_2	0.139 ± 0.621	0.574	0.969	2.06	0.115 ± 0.402	0.356	0.548	1.65
$(A_1)_{11}$	0.326 ± 0.391	0.404	0.312	1.29	0.368 ± 0.250	0.256	0.215	1.19
$(A_1)_{21}$	0.223 ± 0.475	0.564	0.370	1.57	0.260 ± 0.305	0.347	0.258	1.38
$(A_1)_{12}$	1.152 ± 0.329	0.380	0.261	1.51	1.175 ± 0.215	0.236	0.185	1.32
$(A_1)_{22}$	0.629 ± 0.407	0.437	0.310	1.30	0.667 ± 0.263	0.280	0.231	1.21
$(A_2)_{11}$	0.353 ± 0.292	0.305	0.231	1.42	0.351 ± 0.191	0.205	0.162	1.33
$(A_2)_{21}$	-0.402 ± 0.356	0.320	0.372	1.47	-0.397 ± 0.234	0.210	0.250	1.35
$(A_2)_{12}$	-0.206 ± 0.443	0.283	0.543	1.51	-0.256 ± 0.283	0.207	0.340	1.39
$(A_2)_{22}$	-0.418 ± 0.543	0.374	0.640	1.45	-0.460 ± 0.347	0.270	0.398	1.31
	$N = 100$				$N = 400$			
w_1	0.256 ± 0.224	0.198	0.286	1.46	0.252 ± 0.110	0.099	0.123	1.20
w_2	0.107 ± 0.274	0.247	0.327	1.37	0.104 ± 0.134	0.125	0.146	1.17
$(A_1)_{11}$	0.384 ± 0.168	0.170	0.152	1.13	0.396 ± 0.082	0.082	0.077	1.06
$(A_1)_{21}$	0.281 ± 0.206	0.227	0.180	1.27	0.295 ± 0.100	0.106	0.092	1.14
$(A_1)_{12}$	1.188 ± 0.146	0.160	0.125	1.24	1.197 ± 0.071	0.077	0.067	1.15
$(A_1)_{22}$	0.682 ± 0.178	0.182	0.160	1.11	0.695 ± 0.087	0.092	0.082	1.10
$(A_2)_{11}$	0.352 ± 0.131	0.142	0.114	1.26	0.350 ± 0.064	0.066	0.060	1.11
$(A_2)_{21}$	-0.401 ± 0.159	0.150	0.166	1.23	-0.400 ± 0.078	0.077	0.081	1.13
$(A_2)_{12}$	-0.276 ± 0.192	0.149	0.215	1.24	-0.294 ± 0.093	0.085	0.101	1.15
$(A_2)_{22}$	-0.479 ± 0.234	0.199	0.264	1.24	-0.494 ± 0.113	0.102	0.124	1.15

The simulation results in Table 1 show that the least squares estimates of the AR parameters are biased when the sample size is small [cf. Tjøstheim and Paulsen 1983; Mentz et al. 1998]. Consistent with asymptotic theoretical results on the bias of AR parameter estimates [Tjøstheim and Paulsen 1983], the bias of the least squares estimates in the simulations decreases roughly as $1/N$ as the sample size N increases. The bias of the estimates affects the reliability of the confidence intervals because in the approximate confidence intervals $\hat{\phi} \pm \hat{\phi}_\pm$, centered on the least squares estimate $\hat{\phi}$, the bias is not taken into account. The bias of the estimates is one of the reasons why, for each parameter ϕ, the median of the approximate 95% margins of error $\hat{\phi}_\pm$ differs from the absolute values ϕ_- and ϕ_+ of the 2.5th percentile and of the 97.5th percentile of the simulated estimation error distribution [cf. Nankervis and Savin 1988]. For small sample sizes N, the absolute values ϕ_- and ϕ_+ of the 2.5th percentile and of the 97.5th percentile of the estimation error distribution differ considerably, for $N = 25$ by nearly a factor of two. Nevertheless, the median of the approximate margins of error $\hat{\phi}_\pm$ lies, for each parameter ϕ, in between the absolute values of the percentiles ϕ_- and ϕ_+ of the simulated estimation error distribution. The approximate margins of error $\hat{\phi}_\pm$ can thus be used as rough indicators of the magnitudes of the estimation errors. But as the values of q_{95} suggest, the approximate margins of error are reliable indicators of the magnitudes of the estimation errors only when the sample size N is large.

We carried out an analogous analysis for the eigendecomposition of the estimated

AR(2) models. Imposing the normalization conditions (21) on the eigenvectors

$$\tilde{S}_{:k} = \begin{pmatrix} \lambda_k S_{:k} \\ S_{:k} \end{pmatrix}$$

of the augmented coefficient matrix

$$\tilde{A} = \begin{pmatrix} A_1 & A_2 \\ I & 0 \end{pmatrix}$$

that consists of the above coefficient matrices (46), one finds, for the simulated process (44), the eigenmodes

$$S_{:1} = \begin{pmatrix} 0.750 \\ -0.301 \end{pmatrix}, \qquad S_{:2} = \begin{pmatrix} 0.768 \\ -0.362 \end{pmatrix}, \qquad S_{:3,4} = \begin{pmatrix} 0.495 \pm 0.315i \\ 0.323 \mp 0.397i \end{pmatrix}, \quad (48)$$

and the eigenvalues

$$\lambda_1 = -0.728, \qquad \lambda_2 = 0.623, \qquad \lambda_{3,4} = 0.603 \pm 0.536i. \quad (49)$$

Associated with the eigenmodes are the periods

$$T_1 = 2, \qquad T_2 \to \infty, \qquad T_{3,4} = 8.643, \quad (50)$$

and the damping times

$$\tau_1 = 3.152, \qquad \tau_2 = 2.114, \qquad \tau_{3,4} = 4.647. \quad (51)$$

The eigenmodes, periods, and damping times obtained from the ensembles of estimated models were compared with the eigenmodes, periods, and damping times of the simulated process.

Table 2 and Table 3 show, for functions $\phi(x_B)$ of the AR parameters B, the median of the estimates $\hat{\phi} = \phi(x_{\hat{B}})$ and the median of the margins of error $\hat{\phi}_\pm$ belonging to the approximate 95% confidence intervals (40). The function ϕ stands for a real part $\operatorname{Re} S_{jk}$ or an imaginary part $\operatorname{Im} S_{jk}$ of a component S_{jk} of an eigenmode, or for a period T_k or a damping time τ_k. As in Table 1, the symbols ϕ_- and ϕ_+ refer to the absolute values of the 2.5th percentile and of the 97.5th percentile of the simulated distribution of the estimation errors $\hat{\phi} - \phi$, and the quantity q_{95} is defined as above. A value of "NaN" for the quantity q_{95} stands for the indefinite expression 0/0. A value of infinity ("Inf") results from the division of a nonzero number by zero.

Which eigenmode, period, and damping time of an estimated AR(2) model corresponds to which eigenmode, period, and damping time of the simulated AR(2) process (44) is not always obvious, in particular not when the effective time series length N is so small that the estimated parameters are affected by large uncertainties. To establish the statistics in Tables 2 and 3, we matched the estimated eigenvalues $\hat{\lambda}_k$ with the eigenvalues λ_k of the simulated process (49) by finding the permutation π of the indices $1, \ldots, 4$ that minimized

$$\sum_{k=1}^{4} |\hat{\lambda}_{\pi_k} - \lambda_k|^2.$$

The estimated eigenmodes, periods, and damping times were matched with the eigenmodes, periods, and damping times of the simulated process according to the

Table 2. Estimates and 95% margins of error for the periods and damping times of the bivariate AR(2) model (44).

ϕ	$\hat{\phi} \pm \hat{\phi}_{\pm}$	ϕ_-	ϕ_+	q_{95}	$\hat{\phi} \pm \hat{\phi}_{\pm}$	ϕ_-	ϕ_+	q_{95}
	$N = 25$				$N = 50$			
T_1	2.000 ± 0.000	0.000	0.000	NaN	2.000 ± 0.000	0.000	0.000	NaN
τ_1	3.285 ± 4.738	2.181	8.677	4.51	3.202 ± 2.904	1.752	4.229	2.89
T_2	$\mathrm{Inf} \pm 0.000$	0.000	0.000	NaN	$\mathrm{Inf} \pm 0.000$	0.000	0.000	NaN
τ_2	1.309 ± 2.491	1.841	4.901	4.26	1.694 ± 2.140	1.646	3.121	3.38
$T_{3,4}$	8.578 ± 2.791	2.553	4.848	3.23	8.650 ± 2.187	1.883	3.420	2.59
$\tau_{3,4}$	6.449 ± 9.807	2.557	21.942	5.43	5.484 ± 5.172	2.112	8.204	2.97
	$N = 100$				$N = 400$			
T_1	2.000 ± 0.000	0.000	0.000	NaN	2.000 ± 0.000	0.000	0.000	NaN
τ_1	3.147 ± 1.922	1.402	2.638	2.22	3.143 ± 0.928	0.802	1.047	1.44
T_2	$\mathrm{Inf} \pm 0.000$	0.000	0.000	NaN	$\mathrm{Inf} \pm 0.000$	0.000	0.000	NaN
τ_2	1.903 ± 1.687	1.325	2.184	2.52	2.071 ± 0.892	0.797	0.970	1.51
$T_{3,4}$	8.643 ± 1.644	1.436	2.405	2.18	8.646 ± 0.889	0.798	1.049	1.51
$\tau_{3,4}$	5.097 ± 3.123	1.698	4.153	2.10	4.747 ± 1.339	1.049	1.634	1.54

Table 3. Estimated eigenmodes (48).

ϕ	$\hat{\phi} \pm \hat{\phi}_{\pm}$	ϕ_-	ϕ_+	q_{95}	$\hat{\phi} \pm \hat{\phi}_{\pm}$	ϕ_-	ϕ_+	q_{95}
	$N = 25$				$N = 50$			
$\mathrm{Re}\, S_{11}$	0.738 ± 0.162	0.139	0.149	1.44	0.745 ± 0.108	0.090	0.115	1.41
$\mathrm{Re}\, S_{21}$	-0.302 ± 0.256	0.252	0.348	2.45	-0.300 ± 0.162	0.158	0.197	1.87
$\mathrm{Re}\, S_{12}$	0.701 ± 0.488	0.926	0.065	12.78	0.742 ± 0.248	0.626	0.042	15.37
$\mathrm{Re}\, S_{22}$	-0.557 ± 0.663	0.628	0.426	3.46	-0.460 ± 0.528	0.606	0.308	2.39
$\mathrm{Re}\, S_{13,4}$	0.477 ± 0.231	0.566	0.176	4.48	0.486 ± 0.195	0.433	0.156	3.97
$\lvert \mathrm{Im}\, S_{13,4} \rvert$	0.329 ± 0.189	0.496	0.194	6.32	0.324 ± 0.184	0.420	0.196	5.19
$\mathrm{Re}\, S_{23,4}$	0.348 ± 0.259	0.628	0.293	5.62	0.338 ± 0.242	0.519	0.264	4.61
$\lvert \mathrm{Im}\, S_{23,4} \rvert$	0.300 ± 0.256	0.518	0.109	3.71	0.336 ± 0.208	0.405	0.119	3.18
	$N = 100$				$N = 400$			
$\mathrm{Re}\, S_{11}$	0.749 ± 0.074	0.066	0.082	1.34	0.750 ± 0.037	0.034	0.038	1.16
$\mathrm{Re}\, S_{21}$	-0.300 ± 0.109	0.110	0.126	1.60	-0.301 ± 0.053	0.050	0.056	1.25
$\mathrm{Re}\, S_{12}$	0.754 ± 0.137	0.367	0.033	13.91	0.766 ± 0.053	0.110	0.023	6.82
$\mathrm{Re}\, S_{22}$	-0.413 ± 0.404	0.506	0.255	1.99	-0.371 ± 0.211	0.256	0.163	1.57
$\mathrm{Re}\, S_{13,4}$	0.492 ± 0.162	0.307	0.135	3.38	0.494 ± 0.104	0.136	0.093	1.97
$\lvert \mathrm{Im}\, S_{13,4} \rvert$	0.320 ± 0.167	0.319	0.183	4.01	0.316 ± 0.115	0.144	0.117	2.03
$\mathrm{Re}\, S_{23,4}$	0.331 ± 0.212	0.377	0.236	3.46	0.324 ± 0.141	0.174	0.139	1.90
$\lvert \mathrm{Im}\, S_{23,4} \rvert$	0.363 ± 0.164	0.287	0.121	2.64	0.388 ± 0.099	0.122	0.089	1.74

minimizing permutation of the indices. To remove the remaining ambiguity of the sign of the estimated eigenmode $\hat{S}_{:\pi_k}$ belonging to the eigenvalue $\hat{\lambda}_{\pi_k}$, we chose the sign of the estimated eigenmode $\hat{S}_{:\pi_k}$ such as to minimize

$$||\hat{S}_{:\pi_k} - S_{:k}||_2.$$

This procedure allowed us to match a parameter $\hat{\phi}$ of the eigendecomposition of an estimated model uniquely with a parameter ϕ of the eigendecomposition of the simulated process.

The estimation results in Tables 2 and 3 show that large samples of time series

data are required to estimate the eigenmodes, periods, and damping rates of an AR model reliably. The approximate 95% margins of error are rough indicators of the estimation errors most of the time, but the approximate margins of error are always reliable as indicators of the magnitude of the estimation error only when the sample size N is large. Even for large samples, the approximate confidence intervals for the eigenmodes, periods, and damping times are less accurate than the confidence intervals for the AR parameters themselves. The median of the approximate margins of error $\hat{\phi}_\pm$ does not always lie in between the percentiles ϕ_- and ϕ_+ of the simulated estimation error distribution. The fact that the absolute values of the percentiles ϕ_- and ϕ_+ in some cases differ significantly even when the bias of the estimates for a parameter ϕ is small indicates that the distribution of the parameter estimates can be skewed, showing that the t ratio (38) does not follow Student's t distribution. The lower accuracy of the confidence intervals for the eigendecomposition compared with the accuray of the confidence intervals for the AR parameters themselves is a consequence of the linearization involved in the construction of the approximate confidence intervals for the eigendecomposition.

6. CONCLUSIONS

If a multivariate time series can be modeled adequately with an AR model, dynamical characteristics of the time series can be examined by structural analyses of the fitted AR model. The eigendecomposition discussed in this paper is a structural analysis of AR models by means of which aspects of a system that oscillate with certain periods and that relax towards a mean state with certain damping times can be identified. Eigendecompositions of AR(1) models have been used in the geosciences to characterize oscillations in complex systems [von Storch and Zwiers 1999, Chapter 15]. We have generalized the eigendecomposition of AR(1) models to AR(p) models of arbitrary order and have shown that the eigendecomposition of higher-order models can be used as a data analysis method in the same way the eigendecomposition of AR(1) models is currently used. Because a larger class of systems can be modeled with higher-order AR(p) models than with AR(1) models, generalizing the eigendecomposition of AR(1) models to AR(p) models of arbitrary order renders this data analysis method more widely applicable.

Since the eigendecomposition of AR models is of interest in particular for high-dimensional data as they occur, for example, when the state vectors of a time series represent spatial data, we have proposed a computationally efficient stepwise least squares algorithm for the estimation of AR parameters from high-dimensional data. In the stepwise least squares algorithm, the least squares estimates for an AR model of order $p < p_{\max}$ are computed by downdating a QR factorization for a model of order $p_{\max}$. The downdating scheme makes the stepwise least squares algorithm a computationally efficient procedure when both the order of an AR model and the AR parameters are to be estimated from large sets of high-dimensional data.

The least squares estimates of the parameters of an AR(p) model are conditional estimates in that in deriving the least squares estimates, p initial state vectors of the available time series data are taken to be constant, although, in fact, they are part of a realization of a stochastic process. Unconditional estimates that are not based on some such approximation are obtained from Gaussian data with exact maximum likelihood procedures [see, e.g., Wei 1994]. By the exact treatment of the p initial

state vectors, the problem of maximizing the likelihood becomes nonlinear, so that exact maximum likelihood algorithms are iterative and usually slower than least squares algorithms [cf. Wei 1994, Chapter 14]. To be sure, with an exact maximum likelihood algorithm such as that of Ansley and Kohn [1983, 1986], stability of the estimated model can be enforced, the time series data can have missing values, and model parameters can be constrained to have given values, which with linear least squares algorithms is impossible. But for high-dimensional data without missing values, the computational efficiency of the stepwise least squares algorithm might be more important than the guarantee that the estimated AR model be stable or satisfy certain constraints on the parameters. The conditional least squares estimates might then be used as initial values for an exact maximum likelihood algorithm. Or, because it appears that for AR models the conditional least squares estimates are of an accuracy comparable with the accuracy of the unconditional maximum likelihood estimates [cf. Mentz et al. 1998], the stepwise least squares algorithm might be used in place of computationally more complex exact maximum likelihood algorithms.

Approximate confidence intervals for the eigenmodes and their periods and damping times can be constructed from the asymptotic distribution of the least squares estimator of the AR parameters. For lack of a distributional theory for the eigenvectors and eigenvalues of Gaussian random matrices, the confidence intervals for the eigenmodes, periods, and damping times are based on linearizations and rough approximations of the distribution of estimation errors. Simulations of a bivariate AR(2) process illustrate the quality of the least squares AR parameter estimates and of the derived estimates of the eigendecomposition of an AR(2) model. The simulations show that the confidence intervals for the eigendecomposition roughly indicate the magnitude of the estimation errors, but that they are reliable only when the sample of available time series data is large.

ACKNOWLEDGMENTS

We thank Dennis Hartmann for introducing us to the principal oscillation pattern analysis; Jeff Anderson, Stefan Dallwig, Susan Fischer, Johann Kim, and Steve Griffies for comments on drafts of this paper; and Heidi Swanson for editing the manuscript.

REFERENCES

Akaike, H. 1971. Autoregressive model fitting for control. *Ann. Inst. Statist. Math. 23*, 163–180.

Anderson, T. W. 1984. *An Introduction to Multivariate Statistical Analysis* (2nd ed.). Series in Probability and Mathematical Statistics. Wiley, New York.

Ansley, C. F. and Kohn, R. 1983. Exact likelihood of a vector autoregressive moving average process with missing or aggregated data. *Biometrika 70*, 275–278.

Ansley, C. F. and Kohn, R. 1986. A note on reparameterising a vector autoregressive moving average model to enforce stationarity. *J. Statist. Comp. Sim. 24*, 99–106.

Björck, Å. 1996. *Numerical Methods for Least Squares Problems*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

Draper, N. and Smith, H. 1981. *Applied Regression Analysis* (2nd ed.). Series in Probability and Mathematical Statistics. Wiley, New York.

Hansen, P. C. 1997. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects*

of Linear Inversion. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics, Philadelphia, PA.

HASSELMANN, K. 1988. PIPs and POPs: The reduction of complex dynamical systems using principal interaction and oscillation patterns. *J. Geophys. Res. 93*, D9, 11015–11021.

HIGHAM, N. J. 1996. *Accuracy and Stability of Numerical Algorithms*. Society for Industrial and Applied Mathematics, Philadelphia, PA.

HONERKAMP, J. 1994. *Stochastic Dynamical Systems: Concepts, Numerical Methods, Data Analysis*. VCH, New York.

HUANG, B. H. AND SHUKLA, J. 1997. Characteristics of the interannual and decadal variability in a general circulation model of the tropical Atlantic Ocean. *J. Phys. Oceanogr. 27*, 1693–1712.

LÜTKEPOHL, H. 1985. Comparison of criteria for estimating the order of a vector autoregressive process. *J. Time Ser. Anal. 6*, 35–52. Correction, *8* (1987), 373.

LÜTKEPOHL, H. 1993. *Introduction to Multiple Time Series Analysis* (2nd ed.). Springer-Verlag, Berlin.

MENTZ, R. P., MORETTIN, P. A., AND TOLOI, C. M. C. 1998. On residual variance estimation in autoregressive models. *J. Time Series Anal. 19*, 187–208.

NANKERVIS, J. C. AND SAVIN, N. E. 1988. The Student's t approximation in a stationary first order autoregressive model. *Econometrica 56*, 119–145.

NEUMAIER, A. 1998. Solving ill-conditioned and singular linear systems: A tutorial on regularization. *SIAM Rev. 40*, 636–666.

PAULSEN, J. AND TJØSTHEIM, D. 1985. On the estimation of residual variance and order in autoregressive time series. *J. Roy. Statist. Soc. B 47*, 216–228.

PENLAND, C. AND SARDESHMUKH, P. D. 1995. The optimal growth of tropical sea surface temperature anomalies. *J. Climate 8*, 1999–2024.

SCHNEIDER, T. AND NEUMAIER, A. 2000. Algorithm 808: ARfit — A Matlab package for the estimation of parameters and eigenmodes of multivariate autoregressive models. *ACM Trans. Math. Softw. 27*, 58–65.

SCHWARZ, G. 1978. Estimating the dimension of a model. *Ann. Statist. 6*, 461–464.

TIAO, G. C. AND BOX, G. E. P. 1981. Modeling multiple time series with applications. *J. Amer. Statist. Assoc. 76*, 802–816.

TJØSTHEIM, D. AND PAULSEN, J. 1983. Bias of some commonly used time series estimates. *Biometrika 70*, 389–399.

VON STORCH, H., BÜRGER, G., SCHNUR, R., AND VON STORCH, J.-S. 1995. Principal oscillation patterns: A review. *J. Climate 8*, 377–400.

VON STORCH, H. AND ZWIERS, F. W. 1999. *Statistical Analysis in Climate Research*. Cambridge University Press, Cambridge, UK.

WEI, W. W. S. 1994. *Time Series Analysis*. Addison-Wesley, Redwood City, CA.

WILKINSON, J. H. 1965. *The Algebraic Eigenvalue Problem*. Monographs on Numerical Analysis. Clarendon Press, Oxford.

XU, J. S. AND VON STORCH, H. 1990. Predicting the state of the Southern Oscillation using principal oscillation pattern analysis. *J. Climate 3*, 1316–1329.

APPENDIX

For the construction of approximate confidence intervals for the eigenmodes, periods, and damping times, one needs the gradient of these functions $\phi(x_B)$ of the AR parameters x_B. The eigendecomposition of an AR model depends only on the AR coefficient matrices $A_1, \ldots, A_p$, but not on the intercept vector w, so that the partial derivatives of the eigenmodes, periods, and damping times with respect to components of the intercept vector w are zero. Because the normalization conditions (10) for the eigenmodes $S_{:k}$ of AR(1) models have the same form as the normalization conditions (21) for the augmented eigenmodes $\tilde{S}_{:k}$ of higher-order

AR(p) models, it suffices to obtain closed-form expressions for the partial derivatives of the eigenmodes $S_{:k}$, periods T_k, and damping times τ_k of AR(1) models. From the partial derivatives for the AR(1) model, the corresponding partial derivatives for higher-order AR(p) models are then obtained by replacing the coefficient matrix A and its eigenvectors $S_{:k}$ by the augmented coefficient matrix $\tilde{A}$ and its eigenvectors $\tilde{S}_{:k}$.

One obtains the partial derivatives of the eigenvectors $S_{:k}$ and of the eigenvalues λ_k by differentiating the normalization conditions (10) and the eigenvector-eigenvalue relation

$$AS_{:k} = \lambda_k S_{:k}.$$

Taking the derivatives leads to the system of equations

$$\begin{aligned} A\dot{S}_{:k} + \dot{A}S_{:k} &= \dot{\lambda}_k S_{:k} + \lambda_k \dot{S}_{:k}, \\ X_{:k}^T \dot{X}_{:k} + Y_{:k}^T \dot{Y}_{:k} &= 0, \\ X_{:k}^T \dot{Y}_{:k} + Y_{:k}^T \dot{X}_{:k} &= 0, \end{aligned}$$

with dotted quantities denoting partial derivatives with respect to an element of the coefficient matrix A. Upon substitution of the eigendecomposition of the coefficient matrix $A = S\Lambda S^{-1}$, these equations become

$$(\Lambda - \lambda_k I)S^{-1}\dot{S}_{:k} - e^{(k)}\dot{\lambda}_k = -S^{-1}\dot{A}S_{:k}, \tag{52}$$

$$X_{:k}^T \dot{X}_{:k} + Y_{:k}^T \dot{Y}_{:k} = 0, \tag{53}$$

$$X_{:k}^T \dot{Y}_{:k} + Y_{:k}^T \dot{X}_{:k} = 0, \tag{54}$$

where $e^{(k)}$ is the kth column of an identity matrix. The kth component of the differentiated eigenvector-eigenvalue relation (52) yields

$$\dot{\lambda}_k = (S^{-1}\dot{A}S)_{kk} \tag{55}$$

as an explicit formula for the partial derivative of the eigenvalue λ_k with respect to the AR coefficients.

If we write the derivative of the eigenvector $S_{:k}$ as

$$\dot{S}_{:k} = SZ_{:k}, \tag{56}$$

the remaining components of the equations (52–54) take the form

$$(\lambda_j - \lambda_k)Z_{jk} = -(S^{-1}\dot{A}S)_{jk} \quad \text{for} \quad j \neq k,$$

$$(X_{:k}^T X + Y_{:k}^T Y)\operatorname{Re} Z_{:k} + (Y_{:k}^T X - X_{:k}^T Y)\operatorname{Im} Z_{:k} = 0,$$

$$(X_{:k}^T Y + Y_{:k}^T X)\operatorname{Re} Z_{:k} + (X_{:k}^T X - Y_{:k}^T Y)\operatorname{Im} Z_{:k} = 0.$$

If all eigenvalues are distinct, these equations can be solved for the matrix Z and yield

$$Z_{jk} = \frac{(S^{-1}\dot{A}S)_{jk}}{\lambda_k - \lambda_j} \quad \text{for} \quad j \neq k, \tag{57}$$

and

$$\operatorname{Re} Z_{kk} = \textstyle\sum_{l \neq k} \left((X^T Y - Y^T X)_{kl} \operatorname{Im} Z_{lk} - (X^T X + Y^T Y)_{kl} \operatorname{Re} Z_{lk} \right), \tag{58}$$

$$\operatorname{Im} Z_{kk} = \frac{\sum_{l\neq k} \left((Y^TY - X^TX)_{kl} \operatorname{Im} Z_{lk} - (X^TY + Y^TX)_{kl} \operatorname{Re} Z_{lk} \right)}{(X^TX - Y^TY)_{kk}}. \quad (59)$$

[In deriving (58) and (59), we used the normalization conditions (10) in the form $(X^TX+Y^TY)_{kk} = 1$ and $(X^TY)_{kk} = 0$.] The expressions (57–59) for the elements of the matrix Z together with the relation (56) give explicit formulas for the partial derivatives of the eigenvectors $S_{:k}$ with respect to the AR coefficients.

In the case of multiple eigenvalues, the system of equations (52–54) cannot be solved uniquely for the partial derivatives of the eigenvectors with respect to the AR coefficients. In this case, however, the eigenvectors are not uniquely determined, so that it is not meaningful to give confidence intervals for them.

Writing the eigenvalues as $\lambda_k = a_k + ib_k$ with real parts a_k and imaginary parts b_k, we deduce from the derivative (55) of the eigenvalues that

$$\dot{a}_k = \operatorname{Re}(S^{-1}\dot{A}S)_{kk}, \qquad \dot{b}_k = \operatorname{Im}(S^{-1}\dot{A}S)_{kk}.$$

It can be verified that the derivatives of the damping time scales (8) and of the periods (9) then take the form

$$\dot{\tau}_k = \tau_k^2 \, \frac{a_k\dot{a}_k + b_k\dot{b}_k}{a_k^2 + b_k^2} \quad (60)$$

and

$$\dot{T}_k = -\frac{T_k^2}{2\pi} \operatorname{Im} \frac{\dot{\lambda}_k}{\lambda_k} = \frac{T_k^2}{2\pi} \, \frac{b_k\dot{a}_k - a_k\dot{b}_k}{a_k^2 + b_k^2}. \quad (61)$$

By means of the formulas (55–61), the ARFIT module **armode** assembles the gradients that are required in the variance estimate (43).