1 Introduction

Mathematical models can describe the dynamic behavior of a system as a function of time and exist in various fields, such as fault diagnosis [1, 2], telecommunication transmission [3–5], control system [6, 7] and signal processing [8, 9]. In most cases, it is not easy to analyse the whole system and to construct the mathematical model based on the analytic approach. Therefore, system identification becomes the first choice when people model a system [10–12], and has a wide application in both linear [13] and nonlinear systems [14–17]. For the multiple-input and multiple-output Box-Jenkins model with disturbances, a two-stage identification method was developed by using the residual model of Kalman filter [18]. Separating the parameters to be estimated into a linear part and a nonlinear part is a common idea to solve the nonlinear least squares problems. Based on this method, Gan et al. eliminated the linear parameters through the orthogonal projection and presented a variable projection algorithm for the radial basis function network-based autoregressive with exogenous inputs model [19].

During the past decade, a great deal of attention has been given to multivariable system identification [20, 21] for the reason that many modern industrial processes are multivariable systems [22]. Applying the scalar system identification methods to multivariable systems may give poor performances, because multivariable systems have high-dimensional variables, complicated structures and many uncertain disturbances [23]. It attracts an increasing interest for researchers to explore more valid methods for multivariable systems [24, 25]. When dealing with the multivariable systems with colored noises, the data filtering technique can be applied to reduce the influence of the noise and improve the estimation accuracy [26–28]. In [29], a filtering based multi-innovation extended stochastic gradient algorithm was developed for improving the parameter estimation accuracy of the multivariable system with the moving average noise.

To simplify the identification model for multivariable systems, the hierarchical identification principle provides a new idea to this problem [30, 31]. The hierarchical identification, which is also known as the decomposition identification, is to decompose a large-scale system into several small size subsystems with fewer variables. However, after the decomposition of some multivariable systems, there are some parameter coupled relations between the subsystems. In this case, we link the parameter estimates inside the subsystems and identify each subsystem based on the coupling identification concept [32]. The coupling identification concept can avoid the redundant estimates and improve the computational efficiency for multivariable systems. In this literature, Ding et al. combined the coupling concept and the gradient search to estimate the non-uniformly sampled data filtering technique and the coupling identification concept [34]. In the previous work [35], we applied the hierarchical identification principle to multivariable system and used the coupling concept to solve the redundant computation.

This work studies the recursive identification methods of multivariable equation-error autoregressive moving average systems by using the data filtering technique and the coupling identification concept. In view of the colored noises, we first design a filter according to the model structure. By filtering the input-output data, the multivariable system is divided into a noise model and a filtered system model. Next, the filtered system model is decomposed into several sub-models on the basis of the number of outputs. However, a major challenge encountered in estimating the parameters is that there are some unmeasurable variables inside the noise system and the sub-models. To obtain the feasible algorithm, we establish some auxiliary models and replace the unknown variables with the outputs of these models. Then, a filtering based partially-coupled generalized extended stochastic gradient algorithm is derived by using the coupling concept and the auxiliary model. Compared with the multivariable generalized extended stochastic gradient (M-GESG) algorithm, the F-PC-GESG algorithm gives more accurate parameter estimates under the same noise level. In addition, we
introduce a forgetting factor to improve the performance of the F-PC-GESG algorithm.

The outline of this paper is as follows. The symbols and the identification model for the multivariable equation-error autoregressive moving average system are given in Section 2. Section 3 derives the filtered identification models and presents the F-PC-GESG algorithm based on the data filtering technique and the coupling concept. Section 4 proposes the M-GESG algorithm and gives some comparisons with the F-PC-GESG algorithm. The numerical simulations are shown to verify the effectiveness of the proposed algorithms in Section 5. Section 6 offers some concluding remarks to end the paper.

2 Problem formulation

In this section, we state the identification problems and derive the hierarchical identification model for multivariable equation-error autoregressive moving average systems. Let us start by introducing some symbols used in this paper. "A := X" or "X := A" stands for "A is defined as X"; the superscript T stands for the vector/matrix transpose; the symbol \( I_m \) denotes an identity matrix of appropriate size \((m \times m)\); \( I_{mn} \) stands for an \( m \)-dimensional column vector whose elements are 1; \( I_{m \times n} \) represents a matrix of size \( m \times n \) whose elements are 1; the symbol \( \otimes \) represents the Kronecker product, for example, \( A \otimes B = [a_{ij}] \in \mathbb{R}^{m \times n} \) and \( B = [b_{ij}] \in \mathbb{R}^{p \times q} \), in general, \( A \otimes B \neq B \otimes A \). \( \text{col}[X] \) is defined as a vector formed by all columns of matrix \( X \) and arranged in order, for example, \( X = [x_1, x_2, \ldots, x_n] \in \mathbb{R}^{m \times n} \), \( x_i \in \mathbb{R}^{m(t)} \), \( i = 1, 2, \ldots, n \), \( \text{col}[X] = [x_1^T, x_2^T, \ldots, x_n^T] \in \mathbb{R}^{m \times n} \).

\( \theta(t) \) denotes the estimate of \( \theta \) at time \( t \); the norm of a matrix (or a column vector) \( X \) is defined by \( \| X \|^2 \) := \( tr(X^T X) \).

Consider the following multi-variable equation-error autoregressive moving average system which is also called the multi-variable controlled autoregressive moving average (CARARMA) system:

\[
A(z)y(t) = B(z)u(t) + D(z)v(t), \quad (1)
\]

where \( y(t) := [y_1(t), y_2(t), \ldots, y_m(t)]^T \in \mathbb{R}^m \) refers to the \( m \)-dimensional output vector, \( u(t) := [u_1(t), u_2(t), \ldots, u_r(t)]^T \in \mathbb{R}^r \) denotes the \( r \)-dimensional input vector, \( v(t) := [v_1(t), v_2(t), \ldots, v_n(t)] \in \mathbb{R}^n \) is a white noise vector, \( A(z) \) and \( B(z) \) are matrix polynomials in the unit backward shift operator \( z^{-1} \).

Define an intermediate variable

\[
w(t) := \frac{D(z)}{C(z)} v(t). \quad (2)
\]

Then Equation (5) can be rewritten as a pseudo-linear regression model:

\[
y(t) = \Phi(t) \theta + v(t). \quad (6)
\]

Equation (6) is the identification model for the multivariable CARARMA system in (1).

Remark 1: For the identification model in (6), the parameter vector \( \theta \) contains all the parameters to be estimated. Although some algorithms can estimate the parameter vector \( \theta \), the high-dimensional parameter vector and information matrix result in a heavy computational burden and poor performance of these algorithms. On the other hand, the information matrix \( \Phi(t) \) consists of the unknown intermediate variables \( u(t-j) \) and the noise terms \( v(t-j) \). The objective of this paper is to present a new effective algorithm using the observation data \( u(t) \) and \( y(t) \). By employing the coupling identification concept and the data filtering technique, the new algorithm can deal with the unknown variables and give more accurate parameter estimates for the multivariable CARARMA system in (1). Since this paper focuses on the parameter estimation, we assume that the orders and the initial values are known, that is to say, the orders \( m, r, n_m, n_r, n_c \) and \( n_d \) are known and \( y(t) = 0 \), \( u(t) = 0 \) and \( v(t) = 0 \) for \( t \leq 0 \).

3 The filtering based partially-coupled generalized extended stochastic gradient algorithm

It is worth noting that the system considered in this paper is disturbed by the autoregressive moving average noise (i.e., colored noise). In order to reduce the impact of the noise, we introduce the data filtering technique here to obtain more accurate estimates. The basic idea is to use a filter to filter the input-output data and to decompose the original identification model into two models which contain a filtered system model and a noise model. Then we further divide the system model into a series of subsystems according to the number of
Define the filtered input vector \( u_f(t) \) and the filtered output vector \( y_f(t) \) as

\[
\begin{align*}
    u_f(t) &:= C(z)u(t) \in \mathbb{R}^r, \\
    y_f(t) &:= C(z)y(t) \in \mathbb{R}^m.
\end{align*}
\]

Multiplying both sides of (1) by \( C(z) \) gives

\[
A(z)C(z)y(t) = B(z)C(z)u(t) + D(z)v(t),
\]

or

\[
A(z)y_f(t) = B(z)u_f(t) + D(z)v(t). \tag{7}
\]

Define the noise parameter vector \( d \), the information matrix \( \phi(t) \) and the filtered information vector \( \varphi(t) \):

\[
\begin{align*}
    d &:= [d_1, d_2, \ldots, d_n_d]^T \in \mathbb{R}^{n_d}, \\
    \phi(t) &:= [v(t - 1), v(t - 2), \ldots, v(t - n_d)] \in \mathbb{R}^{m \times n_d}, \\
    \varphi(t) &:= -y_f(t - 1), -y_f(t - 2), \ldots, -y_f(t - n_d), \\
    u_f(t - 1), u_f(t - 2), \ldots, u_f(t - n_b)]^T \in \mathbb{R}^m.
\end{align*}
\]

Then Equation (7) can be modified as

\[
y_f(t) = [I_m - A(z)]y_f(t) + B(z)u_f(t) + D(z)v(t) = \phi(t)d + \theta^T\varphi(t) + v(t). \tag{8}
\]

Let \( \Omega^i \in \mathbb{R}^{1 \times n_d} \) be the \( i \)th row of the information matrix \( \phi(t) \):

\[
\phi(t) := [\phi_1(t), \phi_2(t), \ldots, \phi_m(t)]^T \in \mathbb{R}^{m \times n_d}.
\]

Similarly, let \( \theta_i(t) \in \mathbb{R}^n \) be the \( i \)th column of the parameter vector \( \theta \) and \( y_{f1}(t) \) be the \( i \)th element of the filtered output vector \( y_f(t) \), that is

\[
\begin{align*}
    \theta &:=[\theta_1, \theta_2, \ldots, \theta_m] \in \mathbb{R}^{n \times m}, \\
    y_{f1}(t) &= [y_{f1}(t), y_{f2}(t), \ldots, y_{fm}(t)]^T \in \mathbb{R}^m.
\end{align*}
\]

Equation (8) can be explicitly written as

\[
\begin{pmatrix}
    y_{f1}(t) \\
    y_{f2}(t) \\
    \vdots \\
    y_{fm}(t)
\end{pmatrix} = \begin{pmatrix}
    \phi_1(t) \\
    \phi_2(t) \\
    \vdots \\
    \phi_m(t)
\end{pmatrix} \begin{pmatrix}
    d \\
    \varphi(t)
\end{pmatrix} + \begin{pmatrix}
    v_1(t) \\
    v_2(t) \\
    \vdots \\
    v_m(t)
\end{pmatrix}. \tag{9}
\]

Then we can decompose the filtered system in (9) into \( m \) filtered subsystem identification models

\[
y_{fi}(t) = \phi_i^T(t)d + \theta_i^T\varphi(t) + v_i(t) \quad i = 1, 2, \ldots, m. \tag{10}
\]

Define the subsystem information vector

\[
\psi_i(t) := \begin{pmatrix}
    \phi_i(t) \\
    \varphi(t)
\end{pmatrix} \in \mathbb{R}^{n + n_d}.
\]

The filtered subsystems in (10) can be expressed as

\[
y_{fi}(t) = \psi_i^T(t) \begin{pmatrix}
    d \\
    \theta_i
\end{pmatrix} + v_i(t). \tag{11}
\]

From the definition of \( w(t) \), we can obtain

\[
C(z)u(t) = D(z)v(t). \tag{12}
\]

Rewrite Equation (12) as the following form

\[
w(t) = \Omega(t)c + \phi(t)d + v(t),
\]

where \( \Omega(t) \) is the noise information matrix and \( c \) is the noise parameter vector:

\[
\Omega(t) := [-w(t - 1), -w(t - 2), \ldots, -w(t - n_c)] \in \mathbb{R}^{n \times n_c}, \\
c := [c_1, c_2, \ldots, c_n]^T \in \mathbb{R}^{n_c}.
\]

Define an intermediate variable \( w_{in}(t) := w(t) - \phi(t)d \in \mathbb{R}^m. \) Then another similar formula can be derived for \( w_{in}(t) \):

\[
w_{in}(t) = \Omega(t)c + w(t). \tag{13}
\]

Equations (11) and (13) are the filter hierarchical subsystem identification model and the noise model for the multivariable CARARMA system in (1). Compared with the identification model in (6), the filter identification models divide the system parameters into two parts. Equation (11) contains the parameter vectors \( \theta_i \) and \( d \), and the noise vector \( c \) is in the noise model (13). Note that the filtered subsystem identification model is disturbed by the white noise \( v(t) \), in other word, we transform the origin system in (1) which has colored noise to \( m \) filtered subsystem models with white noise and a noise model.

Based on the filtered subsystem identification model (11) and the noise model (13), using the negative gradient search principle gives

\[
\begin{pmatrix}
    \dot{d}(t) \\
    \dot{\theta_1}(t) \\
    \dot{\theta_2}(t) \\
    \vdots \\
    \dot{\theta_m}(t)
\end{pmatrix} = \begin{pmatrix}
    d(t - 1) \\
    \theta_1(t - 1) \\
    \theta_2(t - 1) \\
    \vdots \\
    \theta_m(t - 1)
\end{pmatrix} + \frac{\pi(t)}{r_1(t)} \begin{pmatrix}
    \psi_i(t) - \psi_i(t) \dot{d}(t - 1) \\
    \psi_i(t) - \psi_i(t) \dot{\theta_1}(t - 1) \\
    \psi_i(t) - \psi_i(t) \dot{\theta_2}(t - 1) \\
    \vdots \\
    \psi_i(t) - \psi_i(t) \dot{\theta_m}(t - 1)
\end{pmatrix}, \tag{14}
\]

\[
r_1(t) = r_{1,i}(t) = 1 + ||\psi_i(t)||^2, \tag{15}
\]

\[
\dot{c}(t) = \dot{c}(t - 1) + \frac{\Omega_i(t)}{r_2(t)}[w_{in}(t) - \Omega(t)c](t - 1), \tag{16}
\]

\[
r_2(t) = r_2(t - 1) + ||\Omega(t)||^2. \tag{17}
\]

Note that the information vector \( \psi_i(t) \) (\( i = 1, 2, \ldots, m \)) contains the unknown variable \( s(t - j) \), and the information matrix \( \Omega(t) \) is made of the unmeasured intermediate variable \( w(t - j) \). Moreover, since the parameter vector \( c \) is unknown, we cannot obtain the filtered input vector \( u_{f1}(t) \) and the filtered output vector \( y_{f1}(t) \). All these imply that the above relations are impossible to generate the estimates \( d(t), \theta_i(t) \) and \( c(t) \) directly. The solution here is to replace the unknown variables \( w(t - j) \) and \( v(t - j) \) with the outputs \( w(t - j) \) and \( v(t - j) \) of the auxiliary models. Then we can obtain the estimates of \( \Omega(t) \) and \( \phi(t) \):

\[
\Omega(t) := [-w(t - 1), -w(t - 2), \ldots, -w(t - n_c)] \in \mathbb{R}^{n \times n_c}, \\
\phi(t) := [\dot{v}(t - 1), \dot{v}(t - 2), \ldots, \dot{v}(t - n_d)] = [\phi_1(t), \phi_2(t), \ldots, \phi_m(t)]^T \in \mathbb{R}^{n \times n_d}.
\]

According to Equation (4), replacing \( \theta \) with its estimate \( \hat{\theta}(t) \) gives

\[
\dot{w}(t) = y(t) - \hat{\theta}^T(t)\varphi(t).
\]

From the definition of the intermediate variable \( w_{in}(t) \), replacing \( w(t) \) and \( d \) with \( w(t - 1) \) and \( d(t - 1) \), we get

\[
\dot{w}_{in}(t) = w(t - 1) - \phi(t)d(t - 1) = y(t) - \theta^T(t - 1)\varphi(t) - \phi(t)d(t - 1).
\]
Use the parameter estimate
\[ \hat{c}(t) = [\hat{c}_1(t), \hat{c}_2(t), \ldots, \hat{c}_{n_c}(t)]^T \in \mathbb{R}^{n_c \times 1} \]
to construct the estimate of \( \hat{C}(z) \):
\[ \hat{C}(t, z) = 1 + \hat{c}_1(t)z^{-1} + \hat{c}_2(t)z^{-2} + \cdots + \hat{c}_{n_c}(t)z^{-n_c}. \]

Here, we use the filter \( \hat{C}(t, z) \) to filter the input vector and the output vector, and obtain the estimates of the filtered input vector \( \hat{u}_i(t) \) and the filtered output vector \( \hat{y}_i(t) \):
\[
\hat{u}_i(t) = \hat{C}(t, z)u(t) = u(t) + \hat{c}_1(t)u(t-1) + \cdots + \hat{c}_{n_c}(t)u(t-n_c) \]
\[
\hat{y}_i(t) = \hat{C}(t, z)y(t) = y(t) + \hat{c}_1(t)y(t-1) + \cdots + \hat{c}_{n_c}(t)y(t-n_c) \]
\[
= [\hat{y}_1(t), \hat{y}_2(t), \ldots, \hat{y}_{m}(t)]^T. \]

Then the estimate of \( \hat{\phi}_i(t) \) can be defined as
\[
\hat{\phi}_i(t) := [-\hat{y}_1(t-1)^T, -\hat{y}_2(t-2)^T, \ldots, -\hat{y}_i(t-n_a)^T, \hat{u}_1(t-1)^T, \hat{u}_2(t-2)^T, \ldots, \hat{u}_i(t-n_a)^T]^T \in \mathbb{R}^{n_a}.
\]

Subsequently, the information vector \( \hat{\psi}_i(t) \) is constructed by using \( \hat{\phi}_i(t) \) and \( \hat{\varphi}_i(t) \):
\[
\hat{\psi}_i(t) = \left[ \hat{\phi}_i(t), \hat{\varphi}_i(t) \right]^T \in \mathbb{R}^{n_a+n_a}. \]

According to Equation (8), replacing \( y_i(t), \phi_i(t) \) and \( \varphi_i(t) \) with their estimates \( \hat{y}_i(t), \hat{\phi}_i(t) \) and \( \hat{\varphi}_i(t) \), we can get the estimate of \( \psi_i(t) \):
\[
\hat{\psi}_i(t) = \left[ \hat{\psi}_i(t), \hat{\varphi}_i(t) \right]^T. \]

Besides the unknown variables involved in (14)–(17), there is still a problem that each subsystem will produce a parameter estimation vector \( \hat{d}_i(t) \) of the same parameter vector \( d \) at each time \( t \). In order to make it clear, we use \( \hat{d}_i(t) \) to represent the estimate of Subsystem \( i \). Then, replace \( \hat{y}_i(t) \) and \( \hat{\psi}_i(t) \) in (14)–(15) with their estimates \( \hat{y}_i(t) \) and \( \hat{\psi}_i(t) \), and replace \( \hat{\omega}_i(t) \) and \( \Omega_i(t) \) in (16)–(17) with their estimates \( \hat{\omega}_i(t) \) and \( \hat{\Omega}_i(t) \). Through these replacement, we modify Equations (14)–(17) as
\[
\begin{aligned}
\left[ \hat{d}_i(t), \hat{\theta}_i(t) \right] &= \left[ \hat{d}_i(t-1), \hat{\theta}_i(t-1) \right] + \hat{\psi}_i(t) \times \left[ \hat{y}_i(t) - \hat{\psi}_i(t)^T \hat{d}_i(t-1) \right], \\
&= \left[ \hat{d}_i(t-1), \hat{\theta}_i(t-1) \right] + \hat{\psi}_i(t) \times \left[ \hat{y}_i(t) - \hat{\psi}_i(t)^T \hat{d}_i(t-1) \right] \\
&+ \hat{\psi}_i(t) \times \left[ \hat{y}_i(t) - \hat{\psi}_i(t)^T \hat{d}_i(t-1) \right],
\end{aligned}
\]
\[
\hat{\theta}_i(t) = \hat{\theta}_i(t-1) + \hat{\psi}_i(t). \]

It is worth noting that there are \( \hat{d}_1(t), \hat{d}_2(t), \ldots, \hat{d}_m(t) \) for \( i = 1, 2, \ldots, m \) in (19)–(22), and this leads to many redundant parameter estimates. However, cutting down the redundant parameter estimates and improving the parameter estimation accuracy are our aims to explore new identification methods. In general, we usually desire that the proposed algorithm is convergent, that is to say, the parameter estimates approach their true values as the time \( t \) increases. Therefore, we can assume that the parameter estimate
7. Form \( \hat{\theta}(t) \) by (41), compute \( \hat{\psi}(t) \) and \( \hat{\varphi}(t) \) using (39)-(40).
8. Increase \( t \) by 1 and go to Step 2.

The flowchart of computing \( \hat{c}(t) \), \( \hat{\theta}(t) \) and \( \hat{d}_m(t) \) in the F-PC-GESG algorithm is shown in Figure 1.

![Flowchart](image)

**Fig. 1:** The flowchart of computing the F-PC-GESG parameter estimates \( \hat{\theta}(t) \), \( \hat{d}_m(t) \) and \( \hat{c}(t) \).

**Remark 2:** From the F-PC-GESG algorithm in (23)-(41), we can see that only part of the parameters are coupled. To be more specific, only \( \hat{d}_i(t) \) are coupled between the filtered subsystems. The parameter vectors \( \hat{\theta}_i(t) \) are independent, because every subsystem has a corresponding \( \hat{\theta}_i(t) \). This is the meaning of the partially-coupled algorithm.

**Remark 3:** To improve the performance of the F-PC-GESG algorithm, we introduce forgetting factors \( \lambda_1 \) and \( \lambda_2 \) to the F-PC-GESG algorithm. Replace (24), (26) and (28) with Equations (42)-(44), and remain other formulas unchanged for the F-PC-GESG algorithm in (23)-(41):

\[
\begin{align*}
    r_{1,i}(t) &= \lambda_1 r_{1,i}(t-1) + \| \hat{\psi}_i(t) \|^2, \quad i = 2, 3, \ldots, m, \quad (42) \\
    r_{1,1}(t) &= \lambda_1 r_{1,1}(t-1) + \| \hat{\psi}_1(t) \|^2, \quad 0 \leq \lambda_1 < 1, \quad (43) \\
    r_2(t) &= \lambda_2 r_2(t-1) + \| \hat{\Phi}(t) \|^2, \quad 0 \leq \lambda_2 < 1. \quad (44)
\end{align*}
\]

Since the F-PC-GESG algorithm produces the data saturation problem with the time length increasing. The forgetting factors can reduce the weight of the past data and improve the estimation accuracy. The forgetting factors \( \lambda_1 \) and \( \lambda_2 \) can be the same or different in (42)-(44).

**Remark 4:** Compared with the algorithms in the previous work [33] and [34], this paper considers on the parameter identification problems for multivariable equation-error system with the autoregressive moving average noises. In order to eliminate the effect of the colored noises, the F-PC-GESG algorithm uses the data filtering technique to transform the system into a filtered system and a noise system. Then we decompose the filtered system into \( m \) subsystems and estimate the parameters by the coupling identification concept.

### 4 The multivariable generalized extended stochastic gradient algorithm

In the following, we discuss the multivariable generalized extended stochastic gradient (M-GESG) algorithm for comparison. As we can see from the identification model (6), the parameter vector \( \theta \) contains parameters \( \theta_i, c_i, (i = 1, 2, \ldots, n_c \) and \( d_i, (i = 1, 2, \ldots, n_d) \) of System (1).

In consideration of the unknown variables \( w(t-j) \) and \( v(t-j) \), we also employ the auxiliary model method as we do in the F-PC-GESG algorithm. Let \( \hat{w}(t-j) \) and \( \hat{v}(t-j) \) be the outputs of the auxiliary models, and define the estimate of \( \psi(t) \):

\[
\begin{align*}
    \hat{\psi}(t) &= \det \begin{bmatrix} -\hat{w}(t-1), -\hat{w}(t-2), \ldots, -\hat{w}(t-n_c) \\
    \hat{v}(t-1), \hat{v}(t-2), \ldots, \hat{v}(t-n_d) \end{bmatrix} \in \mathbb{R}^{m \times (n_c+n_d)}.
\end{align*}
\]

Then, the estimate of the information matrix \( \Phi(t) \) can be constructed by \( \varphi(t) \) and \( \hat{\theta}(t) \):

\[
\begin{align*}
    \Phi(t) &= \begin{bmatrix} \hat{\psi}(t), \hat{\varphi}(t)^T \otimes I_m \end{bmatrix} \in \mathbb{R}^{m \times n_0}.
\end{align*}
\]

Replacing \( \Phi(t), \theta \) and \( \vartheta \) with their estimates \( \hat{\Phi}(t), \hat{\theta}(t) \) and \( \hat{\vartheta}(t) \) in (4) and (6), \( \hat{w}(t) \) and \( \hat{v}(t) \) can be computed by

\[
\begin{align*}
    \hat{w}(t) &= y(t) - \hat{\theta}^T(t) \varphi(t), \\
    \hat{v}(t) &= y(t) - \hat{\Phi}(t) \hat{\theta}(t).
\end{align*}
\]

Replacing \( \Phi(t) \) in (6) with its estimate \( \hat{\Phi}(t) \) and using the negative gradient search principle, we can obtain the following M-GESG algorithm:

\[
\begin{align*}
    \hat{\vartheta}(t) &= \hat{\vartheta}(t-1) + \frac{\hat{\varphi}(t)^T}{r(t)} \left[ y(t) - \Phi(t) \hat{\theta}(t-1) \right], \\
    r(t) &= r(t-1) + |\hat{\Phi}(t)|^2, \\
    \hat{\Phi}(t) &= \begin{bmatrix} \hat{\psi}(t), \hat{\varphi}(t)^T \otimes I_m \end{bmatrix}, \\
    \varphi(t) &= \begin{bmatrix} -y^T(t-1), -y^T(t-2), \ldots, -y^T(t-n_a) \\
    u^T(t-1), u^T(t-2), \ldots, u^T(t-n_b) \end{bmatrix}^T, \\
    \hat{\psi}(t) &= \begin{bmatrix} -\hat{w}(t-1), -\hat{w}(t-2), \ldots, -\hat{w}(t-n_c) \\
    \hat{v}(t-1), \hat{v}(t-2), \ldots, \hat{v}(t-n_d) \end{bmatrix}, \\
    \hat{w}(t) &= y(t) - \hat{\theta}^T(t) \varphi(t), \\
    \hat{v}(t) &= y(t) - \hat{\Phi}(t) \hat{\theta}(t), \\
    \hat{\theta}(t) &= \begin{bmatrix} \hat{\theta}(t) \\
    \text{col} \hat{\theta}^T(t) \end{bmatrix}.
\end{align*}
\]

The procedures involved in the M-GESG algorithm in (45)-(52) are listed as follows.

1. Set the initial values: let \( t = 1, \hat{\theta}(0) = 1_{n_0}/p_0, \ r(0) = 1, \hat{v}(-j) = 1_{m}/p_0, \hat{w}(-j) = 1_{m}/p_0, j > 0, p_0 = 10^9 \).
2. Collect the observation data \( u(t) \) and \( y(t) \), and construct the information vector and matrices \( \phi(t) \), \( \psi(t) \) and \( \Phi(t) \) using (48)–(49) and (47).
3. Compute \( r(t) \) using (46) and update the parameter estimation vector \( \hat{\theta}(t) \) by (45).
4. Read \( \hat{y}(t) \) and \( \hat{\psi}(t) \) from \( \hat{\theta}(t) \) by (52), and compute \( \hat{w}(t) \) and \( \hat{v}(t) \) by (50)–(51).
5. Increase \( t \) by 1 and go to Step 2.

Remark 5: By means of the auxiliary model identification idea, the M-GESG algorithm in (45)–(52) handles the unknown information matrix \( \Phi(t) \) by replacing it with its estimate \( \hat{\Phi}(t) \) in order to guarantee the realization of the algorithm.

Remark 6: Although the M-GESG algorithm in (45)–(52) can produce the parameter estimation vector \( \hat{\theta}(t) \), the weakness is that \( \hat{\Phi}(t) \) is a large dimension informational matrix, which gives rise to heavy computational burden. Differing from the M-GESG algorithm, the F-PC-GESG algorithm in (23)–(41) divides the system in (1) into one noise model and \( m \) filtered subsystems. Moreover, the F-PC-GESG algorithm uses the data filtering to reduce the influence of the noise and improve the estimation accuracy. For every \( t \), the F-PC-GESG algorithm first gets the noise parameter vector \( \epsilon(t) \), and uses it to calculate the filtered input and output, then identifies the parameters \( \theta(t) \) and \( \hat{d}(t) \) based on the coupled relations between these filtered subsystems. This property also can be seen in Figure 1.

Remark 7: The computational efficiency of an algorithm can be measured by using flops. The computational costs of the M-GESG algorithm and the F-PC-GESG algorithm are listed in Tables 1–2, where \( n_0 = n_c + n_d + mn \), \( n_1 = n + n_d \) and \( n = mn_a + mn_b \). In order to make it clear, we take an example. Assume that \( m = 10 \), \( r = 5 \), \( n_0 = 10 \), \( n_1 = 10 \), \( n_c = 10 \) and \( n_d = 10 \). Then we can calculate \( N_1 - N_2 = 167630 - 17811 = 149819 \). It shows that the F-PC-GESG algorithm has a higher computational efficiency than the M-GESG algorithm. For the system with high orders, the advantage of the F-PC-GESG algorithm becomes more obvious.

The proposed algorithms in this paper can be developed to multivariable bilinear systems with colored noises [36–38], and can combine the neural network methods [39] and the kernel methods [40, 41] to study parameter identification of different systems [42–44].

5 Examples

In this section, we give two numerical simulations to show the effectiveness of the proposed algorithm.

Example 1. Consider the following multivariable system with two-input two-output:

\[
A(z)y(t) = B(z)u(t) + D(z)\psi(t),
\]

\[
A(z) = I_2 + \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} z^{-1},
\]

\[
B(z) = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} z^{-1},
\]

\[
C(z) = I_2 + c_1 z^{-1} + c_2 z^{-2} + c_3 z^{-3},
\]

\[
D(z) = I_2 + d_1 z^{-1} + d_2 z^{-2} + d_3 z^{-3},
\]

\[
\hat{\theta} = \begin{bmatrix} \theta^T \\ \rho \end{bmatrix},
\]

To apply the proposed method, it is necessary to collect the input and output data. Here, we take the inputs \( [u_1(t)] \) and \( [u_2(t)] \) as two independent persistent excitation signal sequences with zero mean and unit variances, and \( [v_1(t)] \) and \( [v_2(t)] \) are taken as two white noise sequences with zero mean and variances \( \sigma^2_v = 0.2 \) for \( v_1(t) \) and \( \sigma^2_v = 0.3 \) for \( v_2(t) \). Then, we can compute the output vector \( y(t) = [y_1(t), y_2(t)]^T \) based on the given input signals, the model and the simulation condition. After obtaining the input and output data, apply the M-GESG algorithm in (45)–(52) and the F-PC-GESG algorithm in (23)–(41) to estimate the parameters of this system. In addition, we add a forgetting factor \( \lambda = 0.99 \) to the F-PC-GESG algorithm to estimate the parameters. The parameter estimates and errors are shown in Tables 3–5. The parameter estimation errors \( \delta := ||\hat{\theta}(t) - \theta||/||\theta|| \) versus \( t \) and the parameter estimates versus \( t \) are shown in Figure 2–4.
Example 2. Consider the following multivariable system:

\[
A(z)y(t) = B(z)u(t) + \frac{d(z)}{c(z)}v(t),
\]

\[
A(z) = I_3 + \begin{bmatrix}
    a_1 & a_2 & a_3 \\
    a_4 & a_5 & a_6 \\
    a_7 & a_8 & a_9
\end{bmatrix} z^{-1},
\]

\[
B(z) = \begin{bmatrix}
    b_4 & b_5 & b_6 \\
    b_7 & b_8 & b_9
\end{bmatrix} z^{-1}.
\]

Here, we take the inputs \{u_1(t)\}, \{u_2(t)\} and \{u_3(t)\} as three independent persistent excitation signal sequences with zero mean and unit variances, and \{v_1(t)\}, \{v_2(t)\} and \{v_3(t)\} are taken as three white noise sequences with zero mean and variances \(\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = 0.1^2\) for \(v_1(t)\), \(v_2(t)\) and \(v_3(t)\). We use the M-GESG algorithm and the PC-GESG algorithm to estimate the parameters of this multivariable system, respectively. Since there are 20 parameters, we do not give the table for parameter estimates and errors.
Table 5 The F-PC-GESG parameter estimates and errors ($\lambda = 0.99$)

<table>
<thead>
<tr>
<th>$t$</th>
<th>$c_{11}$</th>
<th>$c_{12}$</th>
<th>$a_{11}$</th>
<th>$a_{12}$</th>
<th>$a_{21}$</th>
<th>$a_{22}$</th>
<th>$b_{11}$</th>
<th>$b_{12}$</th>
<th>$b_{21}$</th>
<th>$b_{22}$</th>
<th>$\delta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>100</td>
<td>0.14862</td>
<td>-0.76037</td>
<td>0.28162</td>
<td>0.79617</td>
<td>-0.67532</td>
<td>0.73238</td>
<td>0.08518</td>
<td>0.13966</td>
<td>0.09120</td>
<td>-0.12830</td>
<td>20.46260</td>
</tr>
<tr>
<td>200</td>
<td>0.13898</td>
<td>-0.77519</td>
<td>0.26953</td>
<td>0.84149</td>
<td>-0.71902</td>
<td>0.82292</td>
<td>0.08582</td>
<td>0.13969</td>
<td>0.09222</td>
<td>-0.12729</td>
<td>14.56687</td>
</tr>
<tr>
<td>500</td>
<td>0.13454</td>
<td>-0.78673</td>
<td>0.23530</td>
<td>0.89464</td>
<td>-0.76236</td>
<td>0.93976</td>
<td>0.08579</td>
<td>0.13949</td>
<td>0.09144</td>
<td>-0.12640</td>
<td>7.24352</td>
</tr>
<tr>
<td>1000</td>
<td>0.13318</td>
<td>-0.78989</td>
<td>0.24296</td>
<td>0.92922</td>
<td>-0.79208</td>
<td>1.02175</td>
<td>0.08569</td>
<td>0.13935</td>
<td>0.09123</td>
<td>-0.12619</td>
<td>2.95373</td>
</tr>
<tr>
<td>2000</td>
<td>0.12958</td>
<td>-0.79009</td>
<td>0.24026</td>
<td>0.93935</td>
<td>-0.79938</td>
<td>1.04802</td>
<td>0.08569</td>
<td>0.13933</td>
<td>0.09124</td>
<td>-0.12615</td>
<td>2.45025</td>
</tr>
<tr>
<td>3000</td>
<td>0.12887</td>
<td>-0.79010</td>
<td>0.24001</td>
<td>0.93959</td>
<td>-0.79996</td>
<td>1.04698</td>
<td>0.08569</td>
<td>0.13923</td>
<td>0.09124</td>
<td>-0.12615</td>
<td>2.05105</td>
</tr>
</tbody>
</table>

True values 0.14000 -0.80000 0.24000 0.94000 -0.80000 1.05000 0.10000 0.15000 0.12000 -0.10000

Here, we use the parameter estimation errors curve to show the performance of the algorithms. The parameter estimation errors versus $t$ are shown in Figure 5.

Fig. 5: The parameter estimation errors versus $t$

From Tables 3–5 and Figures 2–5, we can draw the following conclusions.

1. The parameter estimation errors of the M-GESG and the F-PC-GESG algorithms become smaller with the data length $t$ increasing – see the estimation errors of the last columns in Tables 3–5.
2. Under the same noise level, the PC-GESG algorithm has a higher parameter estimation accuracy than the M-GESG algorithm – see Tables 3–5 and Figures 2–4. This demonstrates that the proposed algorithm is effective.
3. Introducing a forgetting factor can improve the parameter estimation accuracy of the F-PC-GESG algorithm – see Table 5 and Figure 2.

6 Conclusions

In this paper, the parameter estimation problem has been investigated for multivariable CARARMA systems. An F-PC-GESG algorithm is derived by adopting the filtering technique and the coupling concept. In order to reduce the influence of the noise, we construct a filter to filter the input and output data and transform the system into two parts, including a noise model and a filtered system model. In addition, the filtered system model is divided into several subsystems and identified by the coupling concept. According to the computational comparison, the F-PC-GESG algorithm has less computational burden than the M-GESG algorithm. The simulation results indicate that the F-PC-GESG algorithm can generate more accurate parameter estimates. Moreover, introducing the forgetting factor can improve the performance of the F-PC-GESG algorithm. The basic idea of the proposed method in this paper can be used to study the parameter identification problems of other multivariable systems with different structures and disturbance noises.

The identification method presented in this paper can combine the multi-innovation methods [45] and some mathematical skills [46–48] and statistical methods [49–51] can be used to study the performances of parameter estimation algorithms and can be applied to other fields [52–55].

7 Acknowledgments

This work was supported by the National Natural Science Foundation of China (grant no. 61873111), the 111 Project (grant no. B12018) and the National First-Class Discipline Program of Light Industry Technology and Engineering (grant no. LITE2018-26).

8 References


