Advances in Systems Science and Applications (2013) Vol.13 No.3 233-248

Ways of Fusing Different Types of Information and How Systemic Yoyo Model is Applied in Complex Systems Evaluation and Estimation

Xiaojun Duan[1] and Yi Lin[2]

[1] *Department of Mathematics and Systems Science National University of Defense Technology Changsha 410073, PR China*

[2] *Department of Mathematics Slippery Rock University Slippery Rock, PA 16057 USA*

Abstract

Continuing the works in literature[1], we show in this paper how different types of information can be fused together consistently in order to produce accurate evaluations and estimations for complex systems. The theoretical part of this presentation is based on the standard statistical reasoning, while the ending part constructs three case studies in order to validate the main thinking logic and results obtained in literature[1] and in this paper.

It is shown that (1) for linear systems, when fusing data of different types, the weights placed on the data have profound effects on the outcomes and the achieved precisions, meaning that in this case, the unique optimal weight matrix is determined by the precisions of the data (Gauss-Markov Theorem of linear models); (2) for nonlinear models, when fusing heterogeneous sets of data with varied scales of precision, the structure of the weight matrix is no longer uniquely determined by the precisionsbut also related to the degree of model nonlinearity, indicating that the classical Gauss-Markov Theorem of linear models no longer holds true. At the same time, a specific method of determining the optimal weighting factor and the relevant computational method for estimating the parameters are established.

Combined with the process of conserved information applied in systems evaluation, we provide three case studies, including (1) how to quantitatively measure prior knowledge and observational data so that prior knowledge can be considered in obtaining much improved optimal systems evaluations; (2) how to excavate new sources of observational data of processes so that the established models can be validated jointly using process data collected under different test environments and the directly measured information of the specific indices of concern in order to improve the quality of systems evaluation and estimation and to obtain model validation results of better accuracy; and (3) how to more effectively fuse prior knowledge and heterogeneous sets of data. All of these case studies further witness the epistemological validity of the information conservation existing in the systemic recognition process beneath the systems model description, prior knowledge, and observational data, and their transformational relationship, as obtained in literature[1].

Because other than establishing the theory, particular procedures are provided, conclusions of this work can be directly employed in system evaluations and estimations and related works. This work shows how systemic thinking can be practically applied to benefit the efforts of system evaluation and model estimations involved in various engineering projects.
Keywords Systemic yoyo model, data fusion, linear/nonlinear model, parameter estimation, Gauss-Markov Theorem

1 Introduction

In literature[1], after clarifying the relationship between observational quantities and the target indices to be measured, the systemic yoyo model[2] is established for system evaluations and estimations (Fig.1), where the form of the model, observational data, and prior knowledge are the main sources of information useful for the evaluation, estimation, and prediction of the performance index of the system to be measured.

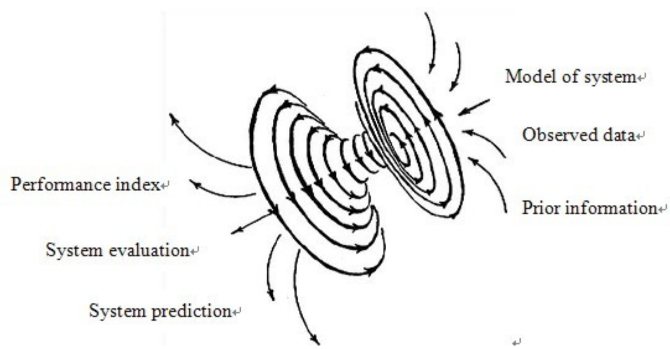

Fig.1 The systemic yoyo model for complex system evaluation

From analyzing the characteristics of and connections between the three main sources of information, model descriptions, prior knowledge, and observational data, for system evaluations and estimations, the following conservation law of information for system analysis is obtained.

$$Ae^{I_M} \times Be^{I_D} \times Ce^{I_P} = a \tag{1}$$

where A, B, and C are constants, I_M stands for the information content described by the model, I_D the information content of the autoptic test data, and I_P the information content of the prior knowledge. The constant a should somehow depict the minimum amount of information required to satisfy the given precision (in the estimate of the model or parameters), where the precision is given in terms of the model accuracy and parameter estimation precision. For the detailed expressions of I_M, I_D, and I_P, see [1].

With this law of conservation is established, Duan and Lin use it to investigate the evolution direction of the process of a system evaluation and estimation.

Continuing what is obtained in literature[1], in this paper, we show that for linear systems, when different types of data are available for systems evaluation and estimation, then the analysis outcomes and precisions achieved are greatly determined by the weights placed on the data. More specifically, the Gauss-Markov Theorem, established on the method of least squares method, holds true. However, when nonlinear systems are involved, the classical Gauss-Markov Theorem of linear models no longer holds. That is, the structure of the weight matrix is not uniquely determined by the precisions of the available data. To this end, we provide a specific method of determining the optimal weighting factor and the relevant computational method for estimating the parameters.

After this theoretical exploration, combined with the conservation law of information of system evaluations and estimations, we construct three case studies to show

(1)How to quantitatively measure prior knowledge and observational data so that better evaluation and estimation results can be obtained using prior knowledge;

(2)How to excavate the available observational data of processes so that the process information collected under different test environments and directly measured data of the specific indices can be employed jointly to fine-tune the model, leading to improved system evaluations, estimations, and more accurate test results, and

(3)How to make prior knowledge and heterogeneous sets of data work effectively together. These case studies further verify the validity of the conservation of information of the model information, prior knowledge, and observational data in the recognition process of systems and the evolutionary relationship between three main sources of information.

This paper is organized as follows: Section 2 looks at various ways one can fuse information in his analysis of complex systems. Section 3 focuses on three specific cases studies. And, the paper is concluded by Section 4.

2 Ways Information Fusion Takes Place in Processes of System Evaluation

With the requirement of precision given, we can optimize the process of a system evaluation. That is such a problem as how to obtain the optimal evaluation results when multiple types of models and multiple kinds of data are available. This end can be analyzed by placing various weights on the multiple kinds of data and prior information.

When dealing with information by combining heterogeneous data, the most typical case is the fusion of such information that are of different types and various precisions. When observational information is expressed by using a parametric model, the problem of how to fuse the information together can be transformed into that of estimating the parameters of some regression models. Here, by different types of information, we mean such information that is of different functional relationships with the parameters to be estimated so that their various orders of derivatives are also different. If in our treatment we have to deal with different types of information of varying precision, different weightings placed on these information will have direct effect on the estimations of the parameters. Sometimes, the effects can also be quite significant. Hence, how to place weights on the information that are of heterogeneous types and varied precisions becomes a key technique for obtaining high accuracies in the parameter estimations.

As for the parameter estimations of linear regression models, Gauss-Markov Theorem provides the most optimal weighting method for observational data of different precisions [3]. As for nonlinear regression models, current publications have assumed that all observational data have the same scale of precision. That is, the random errors in these data are identically independently distributed [3]. To this end, our work in this section theoretically shows that when jointly dealing with heterogeneous sets of data of varied scales of precision, the structure of the weight matrix is no longer uniquely determined by the precision of the observational data; instead, it also has something to do with the degree of nonlinearity of the model, which can be measured by different orders of derivative functions. That is, the classical Gauss-Markov Theorem of linear models does not hold true anymore. In the following, we will specifically address the problem of how to determine the most optimal weighting factor, while providing the relevant computational method for estimating the parameters.

2.1 The Optimally Weighted Information Fusion of Linear System Evaluation

According to the research on the mean square errors of parameter estimations of linear systems, it is readily to show that the estimate corresponding to the optimal weighting factor is the Bayesian estimation, which is Gauss-Markov Theorem[1,6].

For the problem of estimating parameters $\beta_{p\times 1}$, assume that there are the following two types of observational information:

$$\begin{cases} Y_{m\times 1} = X_{m\times p}\beta_{p\times 1} + \varepsilon_{m\times 1} \\ E\varepsilon = 0; Cov(\varepsilon, \varepsilon) = \sigma_1^2 I_{m\times m} \end{cases} \tag{2}$$

and

$$\begin{cases} \widetilde{\beta}_{k\times 1} = Z_{k\times p}\beta_{p\times 1} + \eta_{k\times 1} \\ \eta \sim N(0, \sigma_2^2 I_{k\times k}) \end{cases} \tag{3}$$

satisfying $E\varepsilon\eta^T = 0$.

If we treat model (2) as the direct observational data, while (3) the prior information on the parameters, then the Bayesian estimation of the parameters is the solution of the following extremum problem:

$$\min_{\beta\in R^P} \sigma_1^{-2} \parallel Y - X\beta \parallel_2^2 + \sigma_2^{-2} \parallel \widetilde{\beta} - Z\beta \parallel_2^2 \tag{4}$$

that is given as follows:

$$\widehat{\beta}_B = (\sigma_1^{-2} X^T X + \sigma_2^{-2} Z^T Z)^{-1} (\sigma_1^{-2} X^T Y + \sigma_2^{-2} Z^T \widetilde{\beta}) \tag{5}$$

Therefore, the following conclusions can be shown readily[3-5]:

(1)$E\widehat{\beta}_B = \beta$; that is the Bayesian estimate is unbiased; and

(2)$\mathrm{MSE}(\widehat{\beta}_\mathrm{B}) = \mathrm{tr}(\sigma_1^{-2}\mathrm{X}^\mathrm{T}\mathrm{X} + \sigma_2^{-2}\mathrm{Z}^\mathrm{T}\mathrm{Z})^{-1} < \mathrm{MSE}(\widehat{\beta}_\mathrm{LS}) = \sigma_1^2 \mathrm{tr}(\mathrm{X}^\mathrm{T}\mathrm{X})^{-1}$ that is, by making use of appropriate prior knowledge, one can always improve the estimation precision of the parameters, where $tr(A_{k\times k}) = \sum_{i=1}^k a_{i,i}$ is the trace of the matrix.

When fusing these data, the weights placed on the data have profound effects on the outcomes and the achieved precisions. These conclusions indicate that when fusing observations of different precisions, the unique optimal weight matrix is determined by the precisions of the data, which in essence is still the Gauss-Markov Theorem of linear models established on the least squares method.

2.2 The Optimally Weighted Information Fusion of Nonlinear System Evaluation

For nonlinear models, we show in theory that when fusing heterogeneous sets of data with varied scales of precision, the structure of the weight matrix is no longer uniquely determined by the precisions. That is, the classical Gauss-Markov Theorem of linear models no longer holds true. At the same time, we establish a method for determining the optimal weighting factor and the relevant computational method for estimating the parameters.

For the sake of convenience, for the parameter θ that is to be estimated, we assume that we have the prior information (3) for a linear model, where the error satisfies $E\varepsilon\eta^T = 0$. So, the weighting problem of fusing these two types of observational data can be reduced to the following minimization problem:

$$\min_{\rho\in R^1, \rho>0} \min_{\beta\in R^P} \parallel Y - f(X,\beta) \parallel_2^2 + \rho \parallel \widetilde{\beta} - Z\beta \parallel_2^2 \tag{6}$$

Then, we have the following result:

Theorem1. Denote $S(\beta) = \parallel Y - f(\beta) \parallel_2^2 + \rho \parallel \widetilde{\beta} - Z\beta \parallel_2^2, S(\widehat{\beta}) = \min_\beta S(\beta), C = \sum_{i=1}^m \dot{f}_i^2 + \rho \sum_{i=1}^k z_i^2, D = \sum_{i=1}^m \dot{f}_i \ddot{f}_i, \xi = \sum_{i=1}^m \dot{f}_i \varepsilon_i + \rho \sum_{i=1}^k z_i \eta_i, A = \sigma_1^2 \sum_{i=1}^m \dot{f}_i^2 +$

$\rho^2\sigma_2^2\sum_{i=1}^{k} z_i^2$, then under the assumed conditions (i) and (ii), the following estimation holds true:

$$\begin{aligned}\widehat{\beta}-\beta =& C^{-1}\xi + C^{-2}\sum_{i=1}^{m}\ddot{f}_i\varepsilon_i\xi - \frac{3}{2}C^{-3}D\xi^2 + C^{-3}\sum_{i,j=1}^{m}\ddot{f}_i\ddot{f}_j\varepsilon_i\varepsilon_j\xi \\ & -\frac{9}{2}C^{-4}D\sum_{i=1}^{m}\ddot{f}_i\varepsilon_i\xi^2 + \frac{9}{2}C^{-5}D^2\xi^3\end{aligned} \tag{7}$$

with the bias and mean square error approximated as follows:

$$\begin{cases}\mathrm{E}(\widehat{\beta}-\beta) = -\frac{1}{2}\sigma_1^2\mathrm{C}^{-2}\mathrm{D} - \frac{3}{2}\mathrm{C}^{-3}\mathrm{D}\rho(\rho\sigma_2^2-\sigma_1^2)\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{z}_\mathrm{i}^2 \\ \mathrm{MSE}(\widehat{\beta}) = C^{-2}A + 6C^{-4}D^2\sigma_1^4 + 3C^{-4}A\sigma_1^2\sum_{i=1}^{m}\ddot{f}_i^2 \\ \qquad + \frac{135}{4}C^{-6}D^2A^2 - 36C^{-5}AD^2\sigma_1^2\end{cases} \tag{8}$$

where the assumed conditions (i) and (ii) are given below:

(i) The derivative of $f(t,\beta)$ with respect to the parameter β exists and is continuous, and

$$\lim_{m\to+\infty}\frac{1}{m}\sum_{i=1}^{m} = (\frac{df(t_i,\beta)}{d\beta})^2 = \Omega_1(\beta) > 0 \tag{9}$$

(ii) The second order derivative of $f(t,\beta)$ with respect to β exists and is continuous, and

$$\overline{\lim_{m\to+\infty}}\frac{1}{m}\sum_{i=1}^{m} = (\frac{d^2f(t_i,\beta)}{d\beta^2})^2 = \Omega_2(\beta) \tag{10}$$

Proof. Taking the series expansion of $\dot{S}(\widehat{\beta})$ about the true value of the parameter β produces

$$\dot{S}(\widehat{\beta}) = \dot{S}(\beta) + \ddot{S}(\beta)(\widehat{\beta}-\beta) + 2^{-1}\dddot{S}(\beta)(\widehat{\beta}-\beta)^2$$

Notice that $\dot{S}(\widehat{\beta}) = 0$, $\dot{S}(\beta) = -2(\sum_{i=1}^{m}\dot{f}_i\varepsilon_i + \rho\sum_{i=1}^{m}z_i\eta_i)$, $\ddot{S}(\beta) = 2C - 2\sum_{i=1}^{m}\ddot{f}_i\varepsilon_i$, $\dddot{S}(\beta) = 6D - 2\sum_{i=1}^{m}\dddot{f}_i\varepsilon_i$, So, ignoring all terms of third or higher order derivatives produces

$$\begin{aligned}\widehat{\beta}-\beta &= C^{-1}\Big\{\xi + \sum_{i=1}^{m}\ddot{f}_i\varepsilon_i(\widehat{\beta}-\beta) - \frac{3}{2}D(\widehat{\beta}-\beta)^2 + \frac{1}{2}\sum_{i=1}^{m}\dddot{f}_i\varepsilon_i(\widehat{\beta}-\beta)^2\Big\} \\ &= C^{-1}\xi + C^{-2}\sum_{i=1}^{m}\ddot{f}_i\varepsilon_i\xi - \frac{3}{2}C^{-3}D\xi^2 + C^{-3}\sum_{i,j=1}^{m}\ddot{f}_i\ddot{f}_j\varepsilon_i\varepsilon_j\xi \\ &\quad -\frac{9}{2}C^{-4}D\sum_{i=1}^{m}\ddot{f}_i\varepsilon_i\xi^2 + \frac{9}{2}C^{-5}D^2\xi^3\end{aligned}$$

That is equ.(7) holds true.

Because the expected values of normal variables to odd powers are zero and ε and η are independent, we obtain the first equation in equ. (8) by ignoring the error terms of the fourth and higher orders and then calculating the expected value such that

$$\mathrm{E}(\widehat{\beta}-\beta)^2 = \mathrm{C}^{-2}\mathrm{E}\xi^2+3\mathrm{C}^{-4}\mathrm{E}(\sum_{\mathrm{i}=1}^{\mathrm{m}}\ddot{\mathrm{f}}_{\mathrm{i}}\varepsilon_{\mathrm{i}})^2\xi^2+\frac{45}{4}\mathrm{C}^{-6}\mathrm{D}^2\mathrm{E}\xi^4-12\mathrm{C}^{-5}\mathrm{D}\sum_{\mathrm{i}=1}^{\mathrm{m}}\ddot{\mathrm{f}}_{\mathrm{i}}\varepsilon_{\mathrm{i}}\xi^3 \quad (11)$$

Because

$$\mathrm{E}\xi^2 = \sigma_1^2\sum_{\mathrm{i}=1}^{\mathrm{m}}\dot{\mathrm{f}}_{\mathrm{i}}^2+\rho^2\sigma_2^2\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{z}_{\mathrm{i}}^2\hat{=}\mathrm{A},$$

$$\mathrm{E}\xi^4 = 3\sigma_1^4(\sum_{\mathrm{i}=1}^{\mathrm{m}}\dot{\mathrm{f}}_{\mathrm{i}}^2)^2+6\rho^2\sigma_1^2\sigma_2^2\sum_{\mathrm{i}=1}^{\mathrm{m}}\dot{\mathrm{f}}_{\mathrm{i}}^2\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{z}_{\mathrm{i}}^2+3\rho^4\sigma_2^4(\sum_{\mathrm{i}=1}^{\mathrm{m}}\mathrm{z}_{\mathrm{i}}^2)^2 = 3\mathrm{A}^2,$$

$$\begin{aligned}\mathrm{E}(\sum_{\mathrm{i}=1}^{\mathrm{m}}\ddot{\mathrm{f}}_{\mathrm{i}}\varepsilon_{\mathrm{i}})^2\xi^2 &= 2D^2\sigma_1^4+\sigma_1^4\sum_{i=1}^{m}\dot{f}_i^2\sum_{i=1}^{m}\ddot{f}_i^2+\rho^2\sigma_1^2\sigma_2^2\sum_{i=1}^{m}\ddot{f}_i^2\sum_{i=1}^{k}z_i^2\\ &= 2D^2\sigma_1^4+\sigma_1^2A\sum_{i=1}^{m}\ddot{f}_i^2,\end{aligned}$$

$$\mathrm{E}\sum_{\mathrm{i}=1}^{\mathrm{m}}\ddot{\mathrm{f}}_{\mathrm{i}}\varepsilon_{\mathrm{i}}\xi^3 = 3\mathrm{D}\sigma_1^4\sum_{\mathrm{i}=1}^{\mathrm{m}}\dot{\mathrm{f}}_{\mathrm{i}}^2+3\rho^2\sigma_1^2\sigma_2^2\mathrm{D}\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{z}_{\mathrm{i}}^2 = 3\mathrm{DA}\sigma_1^2,$$

substituting these equations into equ.(11) leads to

$$\mathrm{E}(\widehat{\beta}-\beta)^2 = \mathrm{C}^{-2}\mathrm{A}+6\mathrm{C}^{-4}\mathrm{D}^2\sigma_1^4+3\mathrm{C}^{-4}\mathrm{A}\sigma_1^2\sum_{\mathrm{i}=1}^{\mathrm{m}}\ddot{\mathrm{f}}_{\mathrm{i}}^2+\frac{135}{4}\mathrm{C}^{-6}\mathrm{D}^2\mathrm{A}^2-36\mathrm{C}^{-5}\mathrm{AD}^2\sigma_1^2$$

That is the second equation in equ.(8). QED.

Theorem2. For $\mathrm{MSE}(\widehat{\beta})(\rho)$ in Theorem 1, the solution to the following minimization problem

$$\min_{\rho}\mathrm{MSE}(\widehat{\beta})(\rho) \quad (12)$$

exists, satisfying $\min\limits_{\rho}\mathrm{MSE}(\widehat{\beta})(\rho) < \min_{\rho}\mathrm{MSE}(\widehat{\beta})(\frac{\sigma_1^2}{\sigma_2^2})$.

Proof. Because $\lim\limits_{\rho\to+\infty}\mathrm{MSE}(\widehat{\beta})(\rho) = \sigma_2^2(\sum_{\mathrm{i}=1}^{\mathrm{k}}\mathrm{z}_{\mathrm{i}}^2)^{-1}$, $\mathrm{MSE}(\widehat{\beta})(\rho)$ is infinitely differentiable on $[0, +\infty)$, $\mathrm{MSE}(\widehat{\beta})(\rho)$ has its minimum value on $[0, +\infty)$. Because

$$\begin{aligned} \frac{d}{d\rho}\mathrm{MSE}(\widehat{\beta})(\rho) &= (\sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{z}_{\mathrm{i}}^2)(-2\mathrm{C}^{-3}\mathrm{A} + 2\rho \mathrm{C}^{-2}\sigma_2^2 - 24\mathrm{C}^{-5}\mathrm{D}^2\sigma_1^4 \\ &- 12C^{-5}\sigma_1^2 \sum_{i=1}^{m} \ddot{f}_i^2 + 6\rho C^{-4}\sigma_1^2\sigma_2^2 \sum_{i=1}^{m} \ddot{f}_i^2 - \frac{405}{2} C^{-7} D^2 A^2 \\ &+ 135\rho C^{-6} D^2 A \sigma_2^2 + 180 C^{-6} A D^2 \sigma_1^2 - 72\rho C^{-5} D^2 \sigma_1^2 \sigma_2^2) \end{aligned} \tag{13}$$

we have

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\rho}\mathrm{MSE}(\hat{\beta})(0) = -(\sum_{i=1}^{k} z_i^2)(2(\sum_{i=1}^{m} \dot{f}_i^2)^{-2}\sigma_1^2 + \frac{93}{2}(\sum_{i=1}^{m} \dot{f}_i^2)^{-5} D^2 \sigma_1^4 + \\ 12(\sum_{i=1}^{m} \dot{f}_i^2)^{-5} \sum_{i=1}^{m} \dot{f}_i^2 \sum_{i=1}^{m} \ddot{f}_i^2 \sigma_1^4) < 0 \end{aligned}$$

Also, because $\lim\limits_{\rho \to +\infty} \frac{d}{d\rho}\mathrm{MSE}(\widehat{\beta})(\rho) = 0$, and when $\rho \to +\infty$, each term starting from the third in equ.(13) is a higher order infinitesimal when compared to the previous two terms; and when $\rho > \sigma_1^2\sigma_2^{-2}$, the sum of the previous two terms is greater than zero there is $\rho_0 > 0$ so that $\frac{d}{d\rho}\mathrm{MSE}(\widehat{\beta})(\rho) > 0$ when $\rho \in [\rho_0, +\infty)$. Therefore, the solution of $\min\limits_{\rho} \mathrm{MSE}(\widehat{\beta})(\rho)$ satisfies $\widehat{\rho} \in (0, \rho_0)$.

And because

$$\frac{d}{d\rho}\mathrm{MSE}(\widehat{\beta})(\frac{\sigma_1^2}{\sigma_2^2}) = \frac{\sigma_1^4}{2\mathrm{C}^5} \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{z}_{\mathrm{i}}^2 (33\mathrm{D}^2 - 12(\sum_{\mathrm{i}=1}^{\mathrm{m}} \dot{\mathrm{f}}_{\mathrm{i}}^2 + \frac{\sigma_1^2}{\sigma_2^2} \sum_{\mathrm{i}=1}^{\mathrm{k}} \mathrm{z}_{\mathrm{i}}^2) \sum_{\mathrm{i}=1}^{\mathrm{m}} \ddot{\mathrm{f}}_{\mathrm{i}}^2) \neq 0,$$

$\rho = \sigma_1^2\sigma_2^{-2}$ is not a solution of $\min\limits_{\rho} \mathrm{MSE}(\widehat{\beta})(\rho)$. Hence, Theorem 2 is proven. QED.

Remarks:

(1) Theorem 2 indicates that for nonlinear models, because their least squares estimates are generally biased, when fusing observational data of varied scales of precision, the weight matrix obtained by using Gauss-Markov Theorem, which is derived out of the least squares estimation for linear models, is no longer optimal. The optimal weights can be obtained by solving the minimization problem in equ.(12).

(2) If the prior knowledge equ. (3) is a nonlinear model and the first and second order derivatives of the model are the same as those of nonlinear function $f(X, \beta)$, then the optimal weight can be approximated by $\rho = \sigma_1^2\sigma_2^{-2}$.

2.3 The Parameter Estimation of Heterogeneous Data Fusion

When solving problem (6), one can simply follow the following iterative method:

Step 1: For an initial weight value $\rho_0 = \sigma_1^2 \sigma_2^{-2}$, solve the following minimization problem

$$\min_{\beta \in R^1} \| Y - f(X, \beta) \|^2 + \rho_0 \| \widetilde{\beta} - Z\beta \|^2 \tag{14}$$

to obtain its solution $\widehat{\beta}^{(1)}$;

Step 2: Calculate the mean square error $\mathrm{MSE}(\widehat{\beta}^{(1)})(\widehat{\beta}^{(1)}, \beta)$ of the estimated parameter at $\widehat{\beta}^{(1)}$;

Step 3: Solve the minimization problem $\min\limits_{\rho>0} \mathrm{MSE}(\widehat{\beta}^{(1)})(\widehat{\beta}^{(1)}, \beta)$ to obtain ρ_1; and

Step 4: Repeat Steps 1-4 with the initial value ρ_0 replaced by ρ_1 until the estimated value of the parameter becomes stable.

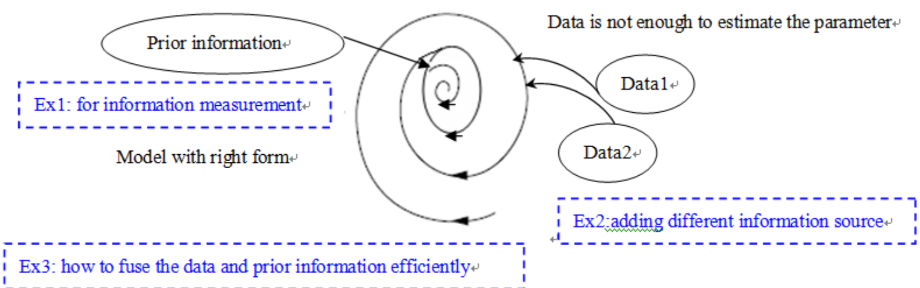

Fig.2 The three samples and their relationships to the information flow of the yoyo model

3 Case Studies

In this section, we will use case studies to illustrate the systemic yoyo model and the evolutionary process naturally existing in system evaluations and estimations. Fig.2 shows the connection of these examples and their individual relationships with the information flow of the systemic yoyo model. We will mainly consider the scenario of supplementing prior information. Because of the shortage of observational data, the convergence of the system evaluation and estimation process is very slow or becomes stagnant after converging to a certain degree. However, after additional prior information becomes available, the stagnated process will continue to converge; and the speed of the resumed convergence is dependent on the quality of the newly supplied prior information and how consistent it is with the observational data.

We will look at three examples to respectively illustrate the following: (1) how to quantitatively measure both the prior and data information; (2) how to excavate a new source of observational process data so that the quality of the ultimate system evaluation and estimation is improved; and (3) how to fuse prior information with heterogeneous sets of data effectively.

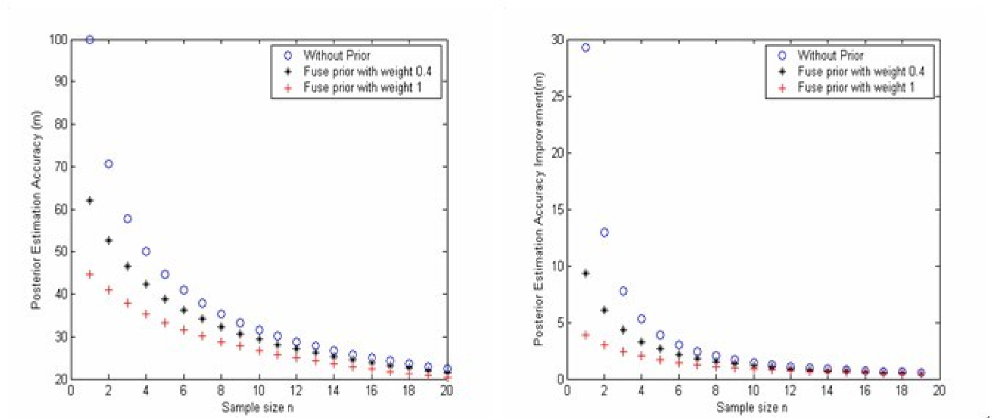

Fig.3 The relationship between the accuracy of post fusion parameter estimation and sample size
(The left represents posterior fusion estimation accuracy (measured by information); the right the posterior fusion accuracy improvement (measured by gain in information))

Table 1 The relation between the Fisher information gain and the increase in sample size

Increase of sample size		1	2	3	4	5	6	7
Information gain	(1)Without prior	29.2893	12.9757	7.7350	5.2786	3.8965	3.0284	2.4411
	(2)Fuse prior,$w = 0.4$	9.3127	6.0794	4.3675	3.3328	2.6511	2.1741	1.8249
	(3)Fuse prior,$w = 1$	3.8965	3.0284	2.4411	2.0220	1.7106	1.4716	1.2836
Increase of sample size		8	9	10	11	12	13	14
Information gain	(1)Without prior	2.0220	1.7106	1.4716	1.2836	1.1325	1.0089	0.9062
	(2)Fuse prior,$w = 0.4$	1.5601	1.3537	1.1892	1.0555	0.9451	0.8527	0.7744
	(3)Fuse prior,$w = 1$	1.1325	1.0089	0.9062	0.8199	0.7464	0.6833	0.6287
Increase of sample size		15	16	17	18	19		
Information gain	(1)Without prior	0.8199	0.7464	0.6833	0.6287	0.5809		
	(2)Fuse prior,$w = 0.4$	0.7075	0.6496	0.5992	0.5551	0.5161		
	(3)Fuse prior,$w = 1$	0.5809	0.5389	0.5017	0.4686	0.4390		

Example1. Let us look at the information measurement of the prior and observational data.
Take the prior parameter variance to be sigma0 = 50 and the sample variance sigma1 = 100. Let us vary the sample size from 1 to 20 and consider three scenarios: no prior information is fused, and prior information is fused with the

consistency weights $w = 1$ and $w = 0.4$, respectively. Figure 3 shows the relationship between the accuracy of the post-fusion parameter estimation and the sample size.

The Fisher information gain is shown in Table 1. Evidently, when the threshold of fusion accuracy is fixed at 1.5, one needs to repeat his test ten times if he does not fuse any additional prior information; if he fuses additional prior information and sets its weight at $w = 0.4$ (that means the consistency between the additional prior information and the test data is measured by weight 0.4), then he needs to repeat the test 9 times; and if he fuses additional prior information with weight set at $w = 1$ (that means the additional prior information has very good consistency with the test data), then he only needs to repeat the test 6 times.

This result indicates that with correct fusion of prior information, the same requirement of precision can be met with a fewer number of times the test is repeated.

Example2. In this case study, we will see how we can speed up the convergence of our recognition of the underlying system by making use of process information and data.

The precision evaluation of active homing radar [6] is a complex recognition process of systems. The impact error of clustered warhead missiles with active homing radar is mainly composed of the measurement error of the Radar navigation system, the instrumental error of the inertial navigation system (INS), the method error of the terminal guidance, the error of the distribution, and some random error. Let us take the impact error of an active homing radar system [6] as our system evaluation performance index and compare the outcomes of the following two methods: one is to fuse the different observational information, directly observed index data, and some indirectly observed data; and the other the point estimate method [7] for the impact error. The main difference here lies in that the later, more traditional method uses only the final impact point error information so that the evaluation outcomes are strongly influenced by the random error of the impact points, while the former method validates the procedure error model with systematic error and the characteristic of the random error by making using of all the observational information. That is how the former method provides more robust evaluation results.

In order to avoid analyzing a heterogeneous population created by different testing states, we will transform these different testing states to standard full range process testing states so that the consequent analysis will become manageable. The impact errors of different testing states are denoted as follows: $\triangle L_{Radar}$ stands for the assessed impact deviation caused by the error of the radar measurement in the standard overall operational test; $\Delta \tilde{L}_{Radar}$ the measured impact deviation caused by the error of Radar measurement in a substitute test.

For a missile with active homing radar, its warhead-target relative positional x-, y-, and z-errors are mainly caused by its radar's measurement error. What an actual radar measures includes the range R, azimuth angle A, and elevation angle E. These measurements satisfy the following transformational connection with the measured warhead-target relative x-, y-, and z-positions:

$$\begin{cases} x = R\cos E\cos A \\ y = R\sin E \\ z = R\cos E\sin A \end{cases} \tag{15}$$

That is, the performance index (that is to be evaluated), the relative positional error of warhead and target, is a function of the observational quantities (range R, azimuth angle A, and elevation angle E). For details, see Fig.4.

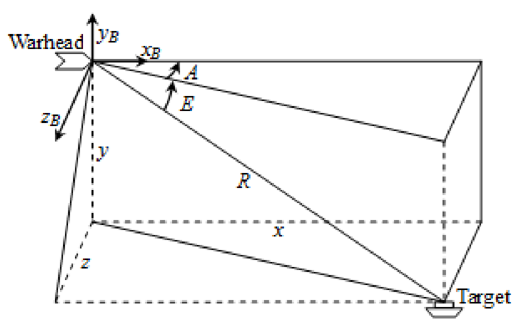

Fig.4 The relationship between radar measurements R, A, and E, and the warhead-target relative x-, y-, and z-positions

According to the Gaussian Law of error propagation, combined with equ. (15), we can obtain the transfer relation from the x-, y-, and z-errors to the R-, A-, and E-errors below:

$$\begin{aligned} \begin{pmatrix} \triangle_x \\ \triangle_y \\ \triangle_z \end{pmatrix} &= \begin{pmatrix} \frac{\partial x}{\partial R} & \frac{\partial x}{\partial A} & \frac{\partial x}{\partial E} \\ \frac{\partial y}{\partial R} & \frac{\partial y}{\partial A} & \frac{\partial y}{\partial E} \\ \frac{\partial z}{\partial R} & \frac{\partial z}{\partial A} & \frac{\partial z}{\partial E} \end{pmatrix} \begin{pmatrix} \triangle_R \\ \triangle_A \\ \triangle_E \end{pmatrix} \\ &= \begin{pmatrix} \cos E\cos A & -R\cos E\sin A & -R\sin E\cos A \\ \sin E & 0 & R\cos E \\ \cos E\sin A & R\cos E\cos A & -R\sin E\sin A \end{pmatrix} \begin{pmatrix} \triangle_R \\ \triangle_A \\ \triangle_E \end{pmatrix} \end{aligned} \tag{16}$$

where $\triangle_x$, $\triangle_y$, and $\triangle_z$ stand for the warhead-target relative positional errors of the standard testing state, and $\triangle_R$, $\triangle_A$, and $\triangle_E$ the errors in the radar measurements R, A, and E of the whole testing state.

Firstly, we use the observational data of actual tests to obtain the errors $\widetilde{\triangle_R}$, $\widetilde{\triangle_A}$, and $\widetilde{\triangle_E}$ of Radar measurements. Secondly, we analyze the iterative process of the errors in the measured range and angles. In the following, we use a Monopulse Radar system as our example to specifically analyze the influencing factors on the errors in radar measured range and angles. Based on the analysis on the sources of errors in radar measurements R, A, and E [6], and the law of error synthesis, we can obtain the main errors in Radar measured ranges and angles as follows:

$$
\begin{aligned}
U_{Angle}^2 &= \triangle_{Rader}^2 + \triangle_{Target}^2 + \triangle_{Enviroment}^2 + \cdots \\
&= \triangle_{ThermaNoise}^2 + \triangle_{PhaseUnbalance}^2 + \triangle_{AngularGlint}^2 + \\
&\quad \triangle_{DynamicLag}^2 + \triangle_{ClutterInterference}^2 + \cdots \\
U_{Distance}^2 &= \triangle_{Rader}^2 + \triangle_{Target}^2 + \triangle_{Enviroment}^2 + \cdots \\
&= \triangle_{ThermaNoise}^2 + \triangle_{AngularGlint}^2 + \triangle_{DynamicLag}^2 + \\
&\quad \triangle_{ClutterInterference}^2 + \cdots
\end{aligned}
$$

If we look at Radar measured ranges, we see that the systematic error is mainly the dynamic lag error. The time-dependent random error mainly includes the error of clutter interference, thermal noise, and distance glint.

Now, let us consider the methods of computation for the impact errors of active homing Radar systems, as mentioned earlier, under two different testing states: One is to fuse different observational information, including the direct observational index data and indirect observational data(the concrete fusion model refers to subsection 2.2 and 2.3), and the other the point estimate method of impact errors [7]. Our simulated Radar measurements are the range R, azimuth angle A, and elevation angle E. Other than analyzing the single point measurements at the impact moments [7], we also provide a method on how to combine observed process quantities.

We respectively model the errors in Radar measured range and angle signals, and the speed and acceleration of the actual measured ranges and angles for the two testing states: the simulated substitute tests and the whole process tests. Assume that the random error term includes the independent errors of clutter interference, thermal noise, and distance glint. For the systematic deviation, we mainly consider the error caused by dynamic lag.

By comparing the point estimate method of impact points and that of combining process information that is used to calculate the impact point error caused by radar errors, the outcomes are listed in Table 2. The point estimate method does not employ the process data of the active homing radar. Instead, it only applies the observational data of the last moments to calculate the impact point deviation caused by Radar errors. When the sample size is small, the outcome

Table 2 The impact error estimate comparisons between the two methods considered (significance level α= 0.01)

Unit (meter)	Cross impact error point estimation	Cross impact error confidence interval	Longitudinal impact error point estimation	Longitudinal impact error confidence interval
Real impact error for standard overall test	-12.93	[-15.72,-10.13]	18.78	[18.09,19.47]
Substitute test point conversion (traditional method)	-13.96	[-49.43,21.51]	20.23	[-29.56,70.02]
Substitute test conversion method fusing with indirect observational information	-12.56	[-15.33,-9.80]	18.33	[17.61,19.05]

of this method is greatly affected by random factors. On the other hand, the method developed in this research righteously employs the physical background information and observed process data so that interval estimates can be directly produced from the estimation of the parameters. Comparing to the point estimate method, our method reduces the effect of random errors and makes the system evaluation and estimation process converge more quickly.

Example3. Let us now look at how to place optimal weights for nonlinear models.

Assume

$$
\begin{aligned}
&f(t,\beta) = 1 + (5 + t\beta)^{0.1},\\
&y(t) = f(t, \beta + \varepsilon(t),\\
&\varepsilon(t) \overset{\text{iid}}{\sim} N(0, 0.01^2),\\
&\widetilde{\beta} = \beta + \eta, \eta \sim N(0, 0.05^2)
\end{aligned}
$$

Let $t = 0.01 * j$, $j = 1, ..., 100$, and the true value of β is 8. Let us generate 50 observational data $\{y^i(t)\}_1^{100}$, $\widetilde{\beta}^i$, i=1,...,50. Then, when $\rho = 0.01^2/0.05^2$, let us solve the minimization problem (6) 50 times, producing the mean square error 0.0394. And when $\rho = 2.33 * 0.01^2/0.05^2$, the mean square error reaches the minimum value 0.0372. However, in theory, the mean square error is 0.0361. This end indicates that for a nonlinear system, the optimal choice of weights for fusing information not only depends on observational precisions but also the model curvature involved.

All the three examples above verify the transformational relationship of system model information, prior knowledge, and observational data in the recognition process of the underlying system, where the model information can be strengthened through the usage of process data, and when the observational data is insufficient, the shortage in information can be made up by supplementing additional prior knowledge.

4 Summary

Continuing literature[1], in this paper we studied how to fuse heterogeneous sets of data together consistently so that better results can be obtained for system evaluations and estimations. It is shown that if the types of data considered are few and the available observational data is insufficient, one can consider obtaining process information and additional prior knowledge. It is because the limited amount of observational data could make the process of system evaluation and estimation converge extremely slowly or stop converging completely after reaching a certain degree. With process information or additional prior knowledge added, the convergence will continue. The speed of the resumed convergence is dependent on how the process information is applied, how good quality the prior knowledge is and how consistent the newly adopted prior information is with the available observational data. This end has been well illustrated by the case studies considered in Section 3. When there is only a small sample available for a specific system evaluation and estimation, this work provides the theoretical guideline for how to excavate other sources of information and how newly adopted information should be fused with what is available.

Acknowledgements

This work is supported by the Natural Science Foundation of China (60974124), the Program for New Century Excellent Talents in University and the Project-sponsored by SRF for ROCS, SEM in China, the key lab open foundation for Space Flight Dynamics technique (SFDLXZ-2010-004).

References

[1] XJ Duan, Y. Lin. (2011), "Conservation law of information and its application in evaluation and estimation of complex systems", *Kybernetes: The International Journal of Cybernetics, Systems, and Management Science*, Vol.40 No.1/2, pp.262-274

[2] Y. Lin. (2008), Systemic yoyos: *Some Impacts of the Second Dimension*, Taylor and Francis, New York.

[3] S. S. Mao. (1999), *Bayesian Statistics.* Beijing: China Statistics Publishing House.

[4] D. M. Bates and D. G. watts. (1997), *Nonlinear Regressive Analysis and Its Application*, Translator: Bocheng Wei. Beijing: China Statistics Publishing House.

[5] X. P. Zhang, J. H. Zhang, and H. W. Xie (2003), "A few discussion of samples, a prior information and Bayesian statistical decision", *Acta Electronica Sinica*, Vol.31 No.4, pp.536-538.

[6] D. C. Wang, J. H. Ding, W. D. Chen (2006), *Radar Measurement Technique in Precise Tracking*, Beijing: Publishing House of Electronics Industry.

[7] G. Wang, X. J. Duan, Z. M. Wang (2009), "Conversion method of impact dispersion in substitute equivalent tests vased on error propagation", *Defence Science Journal*, Vol.59 No.1, pp.15-21.

Corresponding author
Xiaojun Duan can be contracted at: xjduan@nudt.edu.cn