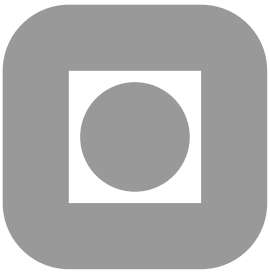
NORGES TEKNISK-NATURVITENSKAPELIGE
UNIVERSITET

Experimental Bayesian Inversion

by

Ayele Taye Goshu and Henning Omre

PREPRINT
STATISTICS NO. 9/2003



NORWEGIAN UNIVERSITY OF SCIENCE AND
TECHNOLOGY
TRONDHEIM, NORWAY

This report has URL http://www.math.ntnu.no/preprint/statistics/2003/S9-2003.ps
Henning Omre has homepage: http://www.math.ntnu.no/∼Omre
E-mail: Omre@stat.ntnu.no
Address: Department of Mathematical Sciences, Norwegian University of Science and Technology,
N-7491 Trondheim, Norway.

Experimental Bayesian Inversion

Ayele Taye Goshu & Henning Omre

Department of Mathematical Sciences,
Norwegian University of Science and Technology,
Trondheim, Norway

Abstract

Experimental inversion is defined as traditional inversion problem with additional complication that the forward model is not directly accessible. It can only be determined including an error term. The usual methods such as Markov chain Monte Carlo (MCMC) sampling and maximum aposteriori (MAP) estimation are not directly available to this extended inverse problem, hence a new experimental Bayesian inversion algorithm is defined. This algorithm is evaluated in a simulation study. It compares favorably with traditional stochastic approximation algorithms in a special case of experimental inversion for which this class of algorithms is applicable. In linear experimental Bayesian inversion problems, the algorithm provides consistent estimates of conditional expectation and conditional variance. The convergence rate appears as faster than the rate for traditional approaches where focus is on estimating the forward model first and thereafter to perform the inversion.

Key Words: inverse problem, experimental model, Bayesian method, algorithm

1 Introduction

Inverse problems appear in many applications, for example image analysis, seismic inversion, signal processing and reservoir characterization. The inverse problems are frequently over or underdetermined and there may be lack of stability in the solution. These characteristics make the inverse problem ill-posed, see Hansen (1998). By casting the inverse problem in a Bayesian setting, a solution may be found in spite of the ill-posedness, see Kolbjørnsen (2002).

The Bayesian formulation includes a likelihood model and a prior model, and the ultimate solution is the posterior model which is uniquely defined by the likelihood and the prior. Let the variable of interest be denoted by x, and assume that a related variable y^o has been observed. The objective of an inverse problem is to characterize x given y^o, termed $[x \mid y^o]$. The likelihood model links the datum y^o to the variable of interest, $y^o = g(x) + u^o$, where $g(x)$ is a known forward function which can be computed for arbitrary values of x, and u^o is an error term which is given a probabilistic interpretation. The likelihood model is termed $f(y^o \mid x)$ and it has the functional form of a probability density function (pdf) but it is a function of the conditioning value x while y^o is a constant. The prior model is represented by a prior pdf on x, i.e. $f(x)$, which often is subjectively assessed. The posterior model is the ultimate

aim of Bayesian inversion and it is represented by the pdf $f(x \mid y^o) = const \times f(y^o \mid x)f(x)$, where *const* is a normalizing constant.

Only in very few cases can the posterior pdf $f(x \mid y^o)$ be analytically determined. The linear-Gaussian case with $g(x)$ linear in x, and u^o and x Gaussian random variables is known to be analytically tractable. Normally, the posterior pdf $f(x \mid y^o)$ will be explored by sampling, often by Markov chain Monte Carlo sampling (MCMC), see Robert and Cassela (1999). Best estimates and estimation variances for $[x \mid y^o]$ can then be obtained from the estimate of $f(x \mid y^o)$. The sampling approach will usually require the forward function $g(x)$ to be computed a very high number of times. In complex inverse problems where x is a spatio-temporal variable for example, even one assessment of $g(x)$ for arbitrary x may require considerable computer resources. Hence sometimes MCMC-sampling is prohibited. In these cases, $[x \mid y^o]$ can be estimated by $[\hat{x} \mid y^o] = \text{argmax}_x\{f(x \mid y^o)\}$, the maximum aposteriori (MAP) estimate, and the estimation variance can be assessed by linearization of $g(x)$ around this estimate and use of Gaussian theory. The MAP-solution can be phrased as an optimization problem, where standard optimization techniques on $g(x)$ can be applied, see Hansen (1998).

The current study is concerned with problems where MCMC-sampling is prohibited from a processing point of view. Hence one must rely on MAP-estimates of $[x \mid y^o]$ and assess the estimation variance by linearization of $g(x)$ around this estimate. An additional complications in the current study is that the forward function $g(x)$ can not be computed exactly for an arbitrary x. One can only obtain $y^* = g(x) + u^*$, where u^* is an error term. This relation is termed an experimental model, and the problem is termed experimental Bayesian inversion. It is obvious that this error term will cause problems in the optimization since repeated runs with the same x as input provide different outputs of y^*.

To the knowledge of the authors, this type of Bayesian inversion problems has not previously been discussed in literature. Further, we feel that this type of inverse problem does exist in practice. One example occurs in a laboratory setting where $g(x)$ only can be assessed through a physical experiment associated with measurement error u^*. Another example appears when $g_*(x_*)$ is a complex computer code requiring a very high-dimensional input x_*, where x_* can be decomposed into $x_* = (x, x_r)$. Let the variable of interest be x, and $[x \mid y^o]$ will be estimated by randomizing over x_r. These examples are further discussed in Section 4.

The problem has some similarity with stochastic approximation as defined in Robbins and Monro (1951) and Frees and Ruppert (1990). In the stochastic approximation setting, there exist one unique solution x to the equation $y^o = g(x)$, and this solution is obtained by sequential sampling of the experimental model. The objective is to determine the unique solution x and asymptotic convergence rate is the criterion. A further discussion on the similarities of experimental Bayesian inversion and the stochastic approximation is included in Section 4.

The rest of the report is organized as follows. In Section 2, the experimental inversion problem is defined. In Section 3, algorithm for the experimental Bayesian inversion is presented. Special cases are discussed and their corresponding algorithms are presented in Section 4. Empirical studies based on three algorithms are considered and their results are discussed in Section 5. An empirical study of the experimental Bayesian inversion algorithm is also presented and discussed in Section 5. Conclusions are forwarded in Section 6. References are

listed in Section 7. Figures are displayed in the Appendix.

2 Problem Definition - Experimental Inversion

The variable $x \in \mathcal{R}^m$ is the variable of interest, and one observation is made as:

$$y^o = g(x) + u^o; \qquad y^o \in \mathcal{R}^1 \tag{1}$$

where $g(.) = \mathcal{R}^m \rightarrow \mathcal{R}^1$ is an unknown function and u^o is an error term including modelling and observation error. Assume further that for a given value of x, one can perform an experiment to obtain:

$$y^* = g(x) + u^*; \qquad y^* \in \mathcal{R}^1 \tag{2}$$

where $g(x)$ is as above and u^* is an experimental error.

The focus of the study is to estimate the value of x given the observed y^o, termed $[x \mid y^o]$, and to quantify the associated estimation uncertainty. More specifically, the objective is to estimate $[x \mid y^o]$ within a pre-specified precision by minimizing the number of runs of the experiment.

This appears as an ill-posed inverse problem with the additional complexity that the forward model can not be exactly reproduced. The problem is termed the experimental inverse problem. It will be cast in a Bayesian setting and the approach will be termed experimental Bayesian inversion.

The stochastic formulation of the problem is as follows:

- Likelihood model

  The observed value y^o is generated through:

  $$[Y^o \mid x] = g(x) + U^o \rightsquigarrow f(y^o \mid x) \rightarrow \mathcal{N}_1(g(x), \sigma_o^2) \tag{3}$$

  with $\mathcal{N}_m(\mu, \Sigma)$ being an m-variate Gaussian pdf with expectation μ and covariance Σ of proper dimensions. The observation likelihood model will then be $f(y^o \mid x)$ as a function of x. Recall however that $g(.)$ is unknown, but σ_o^2 is considered known.

- Prior model

  The variable of interest $x \in \mathcal{R}^m$ is assigned a prior model as follows:

  $$X \rightsquigarrow f(x) \rightarrow \mathcal{N}_m(\mu_X, \Sigma_X) \tag{4}$$

  with parameters (μ_X, Σ_X) considered to be known.

- Experimental model

  The experimental model is defined as follows:

  $$[Y^* \mid x] = g(x) + U^* \rightsquigarrow f(y^* \mid x) \rightarrow \mathcal{N}_1(g(x), \sigma_*^2) \tag{5}$$

  with $g(.)$ unknown as above and σ_*^2 considered known. Moreover, it is assumed that the experiment is expensive to perform.

This formulation is different from standard Bayesian inversion since there is a random component in the experimental model. In Bayesian inversion, one would assume that $g(x)$ is directly obtainable without error for an arbitrary x. The objective in any Bayesian analysis will be to assess the posterior model, and in the Bayesian inversion setting this is $f(x \mid y^o)$. Sampling based inference of the entire posterior pdf is considered to be prohibited since it will require a large number of experiments. In this study, estimating some central location measures of the posterior pdf will be in focus. Moreover, the estimation variance will be assessed through linearization around this estimate. The objective is to determine this estimate within a pre-specified precision through a minimum number of experiments.

The problem above is not frequently discussed in literature, but as mentioned in the introduction, two familiar examples exist. These examples will be briefly discussed below.

Laboratory inversion

Let $y^o \in \mathcal{R}^1$ be a measurement of some phenomenon in the field. The phenomenon is thought of as primarily a simple physical process represented by the unknown $g(x)$ characterized by some parameters. There will, of course, be other less important processes involved and possibly an observation error. These effects are modelled by the error term U^o. The relatively simple physical process can, however, be reproduced in a laboratory where it is controlled by the parameters x. The laboratory experiments are associated with measurement errors U^*. The objective is to assess the parameter value x which gave rise to the field measurement y^o by running the laboratory experiment as few times as possible.

Complex mathematical model inversion

Let $y^o \in \mathcal{R}^1$ be a measurement of some phenomenon in the field. Assume that the phenomenon can be mathematically modelled by a very complex model $g_*(x_*)$ that requires a high number of parameters representd by x_*. It is useful to think about this mathematical model as a computer code requiring x_* as input. The field measurements are associated with error represented by U^o. Assume further that a subset of the parameters $x \subset x_*$ has major influence on $g_*(x_*)$ where $x_* = (x, x_r)$. The objective is to estimate the parameter subset x which gave rise to the measurement y^o, since estimating the full vector x_* is considered to be too difficult.

The inversion defined above can be seen as a slight generalization of the experimental Bayesian inversion model defined earlier, with $x_* = (x, x_r) \in \mathcal{R}^m \times \mathcal{R}^n$:

- Likelihood model

$$[Y^o \mid x] = g(x) + U^o \rightsquigarrow \mathcal{N}_1(g(x), \sigma_o^2)$$

- Prior model

$$X_* \rightsquigarrow \mathcal{N}_{m+n}(\mu_{X_*}, \Sigma_{X_*})$$

- Experimental model

$$[X_r \mid x] \rightsquigarrow \mathcal{N}_n(\mu_{X_r|x}, \Sigma_{X_r|x})$$
$$[Y^* \mid x] = g_*(x, X_r)$$

The unknown function $g(.)$ will then have the interpretation:

$$g(x) = \int g_*(x, x_r) f(x_r \mid x) dx_r = E_{X_r|x}\{g_*(x, X_r)\}$$

which entails that the less important parameters x_r are randomized over.

The objective is to assess the value of the crucial parameters x which gave rise to the field measurement y^o by minimizing the number of calculations of the complex mathematical model.

3 Experimental Bayesian Inversion Algorithm

The experimental Bayesian inversion algorithm is aiming at estimating $[X \mid y^o]$ within a predefined tolerance on the estimation variance by running a minimum number of experiments. The construction of the algorithm is inspired by the following results from linear regression.

Define:

$$[Y^L \mid x] = \dot{x}^T\theta + U^L \rightsquigarrow \mathcal{N}_1(\dot{x}^T\theta, \sigma_L^2)$$

where $x \in \mathcal{R}^m$, $\dot{x} = (1, x)^T$ and $\theta \in \mathcal{R}^{m+1}$ is considered unknown. Consider n experiments with experimental design $(x_1, \ldots, x_n)$ with $n > m + 1$:

$$[Y_1^L \mid x_1], \ldots, [Y_n^L \mid x_n] \quad \text{iid} \quad \mathcal{N}_1(\dot{x}^T\theta, \sigma_L^2)$$

Then from linear regression theory, Mardia et al (1979), the following results are obtained:

$$\begin{aligned}
&\hat{\theta} = (D^T D)^{-1} D^T Y^L \\
&\hat{\theta} \rightsquigarrow \mathcal{N}_{m+1}(\theta, (D^T D)^{-1}\sigma_L^2) \\
&[\hat{Y}^L \mid x^o] = \hat{\theta} x^o \\
&[Y^L \mid x^o] - [\hat{Y}^L \mid x^o] \rightsquigarrow \mathcal{N}_1\left(0, [1 + \dot{x}^{oT}(D^T D)^{-1}\dot{x}^o]\sigma_L^2\right)
\end{aligned}$$

where $D = (\dot{x}_1, \ldots, \dot{x}_n)^T$ and $Y^L = \left(Y_1^L, \ldots, Y_n^L\right)^T$ and $\dot{x}^o = (1, x^o)^T$. Note that $[\hat{Y}^L \mid x^o]$ is the prediction of Y^L at x^o.

Take a closer look at the one-dimensional case, i.e. $m = 1$, with $\theta = (\theta_0, \theta_1)$ and $n \geq 2$,

$$\begin{aligned}
&\hat{\theta} \rightsquigarrow \mathcal{N}_2\left(\theta, \frac{\sigma_L^2}{\sum_{i=1}^n (x_i - \bar{x})^2} \begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}\right) \\
&[Y^L \mid x^o] - [\hat{Y^L} \mid x^o] \rightsquigarrow \mathcal{N}_1\left(0, (\tfrac{1}{n} + 1 + \tfrac{(\bar{x} - x^o)^2}{\sum_{i=1}^n (x_i - \bar{x})^2})\sigma_L^2\right)
\end{aligned}$$

with $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i$.

If estimating the model parameters θ is the objective, minimizing the estimation variance is reasonable. This leads to maximizing $\sum_{i=1}^{n}(x_i - \bar{x})^2$ and hence use a spread experimental design. This is the underlying idea of experimental design theory, see Pukelsheim (1993) and Mayers and Montgomery (2002). If, on the other hand, the objective is to predict Y^L at x^o, minimizing the prediction variance at x^o is reasonable. This leads to setting $\bar{x} - x^o = 0$ which entails centering the experimental design around x^o. This minimizes the prediction variance regardless of the spread of the design. One can, of course, meet both objectives by defining a spread design centred at x^o without increasing the prediction variance. But, if one has some doubt about the assumption of linearity in $E[Y \mid x]$ one may use a narrow design centred at x^o without increasing the prediction variance. This observation is used in the construction of the experimental Bayesian inversion algorithm.

The algorithm draws heavily on results from linear Bayesian inversion, hence some basic results are summarized here. The objective is to estimate $x \in \mathcal{R}^m$ based on the observed value $y^o \in \mathcal{R}^1$. Assume the following model:

- Likelihood model

$$[Y^o \mid x] \rightsquigarrow \mathcal{N}_1(\dot{x}^T\theta, \sigma_o^2) \tag{6}$$

  with $\dot{x} = (1, x)^T$ and $\theta \in \mathcal{R}^{m+1}$ being known.

- Prior model

$$X \rightsquigarrow \mathcal{N}_m(\mu_X, \Sigma_X) \tag{7}$$

  with μ_X and Σ_X being known expectation and covariance matrix of proper dimensions.

- Posterior model

$$[X \mid y^o] \rightsquigarrow \mathcal{N}_m(\mu_{X|y^o}, \Sigma_{X|y^o}) \tag{8}$$

  with

$$\mu_{X|y^o} = E\{X \mid y^o; \theta\} = \mu_X + \Sigma_X\theta\left[\theta^T\Sigma_X\theta + \sigma_o^2\right]^{-1}\left(y^o - \mu_X^T\theta\right) \tag{9}$$

$$\Sigma_{X|y^o} = Var\{X \mid y^o; \theta\} = \Sigma_X - \Sigma_X\theta\left[\theta^T\Sigma_X\theta + \sigma_o^2\right]^{-1}\theta^T\Sigma_X \tag{10}$$

The experimental Bayesian inversion algorithm is a sequential algorithm with the following major steps:

Initiate

- define stop criterion on relative estimation variance
- define initial estimate of $[x \mid y^o]$ termed $\hat{x}_o$

Iterate i=0, 1, 2, ...

- compute new design point x_{i+1} such that $\bar{x}_{i+1} = \frac{1}{i+1}\sum_{j=1}^{i+1} x_j$ is approximately centred at $\hat{x}_i$
- perform experiment at x_{i+1} to obtain y^*_{i+1}
- compute updated estimate, $\hat{x}_{i+1}$
- compute relative estimation variance by bootstrap sampling
- go to STOP if the relative estimation variance is below the stop criterion
- return to iterate

STOP

- Estimate of $[x \mid y^o]$ is $\hat{x}_{i+1}$
- Estimation variance is bootstrap estimate of variance

According to the linear regression theory discussed above, one should define the new design point such that:

$$\bar{x}_{i+1} = \frac{1}{i+1}\sum_{j=1}^{i+1} x_j = \hat{x}_i$$

implying that

$$x_{i+1} = (i+1)\hat{x}_i - i\bar{x}_i$$

since $\hat{x}_i$ is the currently best estimate on $[x \mid y^o]$. Note that x_{i+1} appear as a weighted average of $\hat{x}_i$ and $\bar{x}_i$.

A more general updating procedure, including the one defined above, is used in the algorithm however:

$$\bar{x}_{i+1} = a_i\hat{x}_i + (1-a_i)\bar{x}_i$$

implying that

$$x_{i+1} = a_i(i+1)\hat{x}_i + [1 - a_i(i+1)]\bar{x}_i$$

where $\{a_i\}$ is a sequence of weights. Note that by setting $\{a_i = 1\}$, one obtains the above updating procedure. This choice will later be shown to be unfortunate, however. A reasonable updating procedure is obtained by setting $\{a_i = i^{-\alpha}\}$ with $0.5 < \alpha \leq 1.0$.

The experimental Bayesian inversion algorithm is defined to be:

Algorithm - experimental Bayesian inversion

Initiate

∂ - tolerance in the relative prediction variance $\{a_i\}$ - sequence of weights	A
$[x_1, \ldots, x_{m+1}]$ - initial experimental design	B
$\bar{x}_{m+1} = \frac{1}{m+1} \sum_{j=1}^{m+1} \quad x_j$ $[y_j^* \mid x_j]; j = 1, \ldots, m+1 \rightsquigarrow f(y^* \mid x)$ $V_{m+1}^* = \left(y_1^*, \ldots, y_{m+1}^*\right)^T$ $D_{m+1} = \left(\binom{1}{x_1}, \ldots, \binom{1}{x_{m+1}}\right)^T$ $\hat{\theta}_{m+1} = \left(D_{m+1}^T D_{m+1}\right)^{-1} D_{m+1}^T V_{m+1}^*$ $\hat{x}_{m+1} = E\{X \mid y^o; \hat{\theta}_{m+1}\}$ $\Sigma_{m+1}^{\hat{X}} = Var\{X \mid y^o; \hat{\theta}_{m+1}\}$	C

Iterate $i = m+1, m+2, \ldots$

$x^s \rightsquigarrow \mathcal{N}_{m+1}\left(0, \Sigma_i^{\hat{X}}\right)$ $x_{i+1} = a_i(i+1)\hat{x}_i + [1 - a_i(i+1)]\bar{x}_i + x^s$	D
$[y_{i+1}^* \mid x_{i+1}] \rightsquigarrow f(y^* \mid x)$ Update $\bar{x}_{i+1}, V_{i+1}^*, D_{i+1}$ $\hat{\theta}_{i+1} = (D_{i+1}^T D_{i+1})^{-1} D_{i+1}^T V_{i+1}^*$ $\hat{x}_{i+1} = E\{X \mid y^o; \hat{\theta}_{i+1}\}$ $\Sigma_{i+1}^{\hat{X}} = Var\{X \mid y^o; \hat{\theta}_{i+1}\}$	E
$\Sigma_{i+1}^{\theta} = (D_{i+1}^T D_{i+1})^{-1} \sigma_*^2$ $\theta_l^b; \quad l = 1, \ldots, B \rightsquigarrow \mathcal{N}_{m+1}\left(\hat{\theta}_{i+1}, \Sigma_{i+1}^{\theta}\right)$ $\hat{x}_l^b = E\{X \mid y^o; \theta_l^b\}; \quad l = 1, \ldots, B$ $\Sigma_l^{Xb} = Var\{X \mid y^o; \theta_l^b\}; \quad l = 1, \ldots, B$ $\hat{\Sigma}_{i+1}^X = \frac{1}{B} \sum_{l=1}^B \left(\hat{x}_l^b - \hat{x}_{i+1}\right)\left(\hat{x}_l^b - \hat{x}_{i+1}\right)^T + \frac{1}{B} \sum_{l=1}^B \left(\Sigma_l^{Xb}\right)$	F

if $\left[trace\{\hat{\Sigma}^{X}_{i+1}\}/trace\{\Sigma^{\hat{X}}_{i+1}\}\right] < \partial$ goto STOP | G

STOP

$\hat{x}_{i+1}$ - experimental Bayesian inversion estimate of $[X \mid y^o]$
$\hat{\Sigma}^{X}_{i+1}$ - experimental Bayesian inversion estimate of $Var\{X \mid y^o\}$ | H

This algorithm can be understood as follows:

Block A - Specify a stop criterion for the algorithm. This criterion relates to the bootstrap estimate of $Var\{X \mid y^o\}$ taking the uncertainty in $g(.)$ into account relative to an estimate of $Var\{X \mid y^o\}$ for the true $g(.)$. Specify the weighting sequence $\{a_i\}$ in the selection of design points.

Block B - Specify an initial design of dimension $m+1$ such that a first linear approximation to $g(.)$ can be obtained. This design should be based on classical Bayesian experimental design.

Block C - Perform the initial experiments, estimate a first global linear approximation to $g(.)$ and obtain the initial estimate of $[X \mid y^o]$ with associated estimation variance.

Block D - In iteration $i+1$, determine the new design point x_{i+1} such that the average of the design points is a weighted average of the current estimate of $[X \mid y^o]$ and the average previous design points. An additional centred random term due to the uncertainty in the prior and likelihood models is added. This step focuses the experiments in the area around the solution and robustifies the design against deviations from linearity in $g(.)$.

Block E - In iteration $i+1$, perform the $(i+1)^{th}$ experiment, estimate a new linear approximation to $g(.)$ and obtain a new estimate of $[X \mid y^o]$ with associated estimation variance. Note that the design points are focused such that the linear approximation will be more representative of $g(.)$ around the estimate of $[X \mid y^o]$.

Block F - In iteration $i+1$, provide a bootstrap estimate of $Var\{X \mid y^o\}$ taking also the uncertainty in the approximation to $g(.)$ into account. The following relation is used:

$$Var\{X \mid y^o\} = Var_{\Theta}\{E_X\{X \mid y^o, \Theta\}\} + E_{\Theta}\{Var_X\{X \mid y^o, \Theta\}\}$$

Block G - In iteration $i+1$, terminate iteration when the uncertainty in the approximation of $g(.)$ do not contribute significantly to the estimation of the estimation variance $Var\{X \mid y^o\}$. Note that the best one can obtain is $Var\{X \mid y^o\}$ given the true $g(.)$.

Block H - Stop, use the current values of $\hat{x}_{i+1}$ and $\hat{\Sigma}^{X}_{i+1}$ as the experimental Bayesian inversion estimate of $[X \mid y^o]$ and $Var\{X \mid y^o\}$, respectively.

4 Special Cases

The experimental Bayesian inversion problem has a couple of familiar special cases. These cases are discussed in this section in some detail.

4.1 Traditional Bayesian inversion

Consider the following special case of experimental Bayesian inversion:

- Likelihood model

$$[Y^o \mid x] = g(x) + U^o \rightsquigarrow f(y^o \mid x) \rightarrow \mathcal{N}_1(g(x), \sigma_o^2)$$

- Prior model

$$X \rightsquigarrow f(x) \rightarrow \mathcal{N}_m(\mu_X, \Sigma_X)$$

- Experimental model

$$[Y^* \mid x] = g(x) \rightsquigarrow \delta_1(y^*, g(x))$$

  where $\delta_1(y^*, g(x))$ is the Dirac delta pdf.

Hence this is the experimental Bayesian inversion model with $\sigma_*^2 = 0$, which entails that $g(x)$ can be directly obtained without error for arbitrary x. This corresponds to the traditional Bayesian inversion model, see Kolbjørnsen (2002).

The ultimate objective of Bayesian inversion is to determine the posterior pdf:

$$f(x \mid y^o) = const \times f(y^o \mid x) f(x)$$

If $g(x)$ is linear in x, this posterior can be determined analytically from standard Gaussian theory. When $g(x)$ is non-linear in x, the posterior may be explored by sampling, for example MCMC sampling. In the current case, $g(x)$ is expensive to obtain for an arbitrary x and sampling based inference will be prohibited since it requires a high number of experiments. The usual approach in this situation is to represent $[X \mid y^o]$ by the MAP-estimate:

$$[\hat{X} \mid y^o] = \text{argmax}_x\{f(x \mid y^o)\}$$

which can be determined from an optimization problem. The associated estimation variance is estimated by linearizing $g(x)$ around $[\hat{X} \mid y^o]$ and using standard Gaussian theory.

If the posterior pdf $f(x \mid y^o)$ is multimodal, the optimization problem is hard. Few, if any, algorithms can in practice guarantee global optimization. Simulated annealing claims global optimization properties in theory, but it would require an enormous number of experiments

in practice. The multimodal challenge can be met by assuming that a good initial guess lying in the drainage area of the global optimum can be obtained. In practice, this entails that any optimization algorithm suitable for unimodal posterior pdf $f(x \mid y^o)$ can be applied.

If the posterior pdf $f(x \mid y^o)$ is unimodal, the optimization problem is tractable and numerous algorithms are available. Algorithms that uses gradient information about $g(x)$ in the search for the optimum has proven to be efficient with respect to the number of experiments required. These gradients are, if possible, analytically determined or they are estimated from experiments made in the vicinity of the current estimate of x.

Consider the experimental Bayesian inverse problem as defined above. Recall that the experimental model includes an experimental error term, which entails that there will be an uncertainty associated with every experiment. This will not only make it difficult to identify the optimum whenever it is reached, it will also make estimates of gradients highly unreliable. Hence using standard optimization algorithms in the experimental Bayesian inversion setting is not recommendable. Note further that the gradient estimates are used to assess the estimation variance.

4.2 Stochastic approximation

Consider the following, slightly adapted, special case of experimental Bayesian inversion for $m = 1$, i.e. $x \in \mathcal{R}^1$:

- Likelihood model

$$[Y^o \mid x] = g(x) \rightsquigarrow \delta_1(y^o, g(x))$$

- Prior model

$$X \rightsquigarrow f(x) \rightarrow Uni_1[x_l, x_h]$$

  where $Uni_1[x_l, x_h]$ is the uniform pdf within $[x_l, x_h]$.

- Experimental model

$$[Y^* \mid x] = g(x) + U^* \rightsquigarrow f(y^* \mid x) \rightarrow \mathcal{N}_1(g(x), \sigma_*^2)$$

Hence this is the experimental Bayesian inversion model in one dimension with $\sigma_o^2 = 0$, which entails that the observation is made without error. Moreover, x is assigned a uniform pdf in the interval $[x_l, x_h]$, in stead of being Gaussian, which entails that x is known to be within this interval and that no value of x is more preferred than others a priori. This corresponds to the so called stochastic approximation model in one dimension, see Frees and Ruppert (1990) and Ghosh et al (1997). A multivariate version with $y^o, x \in \mathcal{R}^m$ is also defined, see Wei (1985), but the constraint that y^o and x have to be of same dimension makes it less interesting in practice.

The objective of stochastic approximation is to estimate $x^o = g^{-1}(y^o)$, which has a unique solution only when $g(x)$ is monotone in $[x_l, x_h]$, i.e. $\frac{\partial}{\partial x}g(x) > 0$; $\forall x \in (x_l, x_h)$ or $\frac{\partial}{\partial x}g(x) < 0$; $\forall x \in (x_l, x_h)$. Note, however, that $g(x)$ may be non-linear. The estimate should be obtained by minimizing the number of experiments. Consistency and asymptotic estimation variance as the number of experiments goes to infinity are normally used as criteria.

Two versions of the stochastic approximation algorithm will be presented here. The classical algorithm defined in Robbins and Monro (1951) and a parametric version of it defined in Frees and Ruppert (1990). The latter algorithm has several similarities with the simplified version of the experimental Bayesian inversion algorithm.

Algorithm - classical stochastic approximation algorithm

Initiate

$$x_1 \rightsquigarrow Uni[x_l, x_h]$$
$$[y_1^* \mid x_1] \rightsquigarrow f(y^* \mid x)$$
$$\hat{x}_1 = x_1$$

Iterate $i = 1, 2, \ldots$

$$x_{i+1} = x_i + c_i[y^o - y_i^*]$$
$$[y_{i+1}^* \mid x_{i+1}] \rightsquigarrow f(y^* \mid x)$$
$$\hat{x}_{i+1} = x_{i+1}$$

This defines the so called Robbins-Monro stochastic approximation algorithm, where $\{c_i\}$ is a sequence of tuning constants converging to zero and $\{\hat{x}_i\}$ is a sequence of estimates of x^o.

Under certain relatively weak smoothness assumptions on $g(x)$, it can be shown, see Frees and Ruppert (1990) and Ghosh et al (1997), that if $\{c_i = ci^{-\beta}\}$ with $c > [2g'(x^o)]^{-1}$ and $0 < \beta \leq 1.0$, the asymptotic pdf for the estimator as $i \to \infty$ is:

$$i^{\beta/2}(\hat{x}_i - x^o) \quad \underrightarrow{\mathcal{D}} \quad \begin{cases} \mathcal{N}\left(0, \sigma_*^2 c^2/[2cg'(x^o) - 1]\right) & \text{if} \quad \beta = 1 \\ \mathcal{N}\left(0, \sigma_*^2 c/2g'(x^o)\right) & \text{else} \end{cases}$$

where $g'(x) = \frac{\partial}{\partial x} g(x)$. Moreover, if $\beta = 1.0$ and $c = [g'(x^o)]^{-1}$ it is known that the asymptotic variance is minimized, taking the value $\sigma_*^2 / [g'(x^o)]^2$. The problem is, of course, that $g'(x^o)$ is unknown and has to be estimated itself. In Lai and Robbins (1979), a procedure for this is described, and the algorithm is termed adaptive stochastic approximation. Note that there is an inherent conflict in this algorithm since the objective is to have the estimate sequence $\{\hat{x}_i\}$, which is identical to the sampling sequence $\{x_i\}$, to approach x^o as fast as possible. This makes the sampling design for estimating the slope $g'(x^o)$ unfavorable, however. The crucial statistic is the sampling spread sequence $\{d_i^2\}$ defined by:

$$d_i^2 = \sum_{j=1}^{i} (x_j - \bar{x}_i)^2 \text{ with } \bar{x}_i = \tfrac{1}{i} \sum_{j=1}^{i} x_j$$

Note that d_i^2 corresponds to the denominator in the estimation variance of the slope in the i^{th} iteration. Hence the sequence $\{d_i^2\}$ should grow fast. Under the optimal case with $\beta = 1.0$ and $c = [g'(x^o)]^{-1}$, the growth in $\{d_i^2\}$ is constrained to be:

$$d_i^2 \sim log(i)\sigma_*^2 / [g'(x^o)]^2$$

The alternative stochastic approximation algorithm is termed parametric stochastic approximation algorithm.

Algorithm - parametric stochastic approximation algorithm

Initiate

Set algorithm parameter β

$(x_1, x_2) = (x_l, x_h)$

$[y_j^* \mid x_j]; j = 1, 2 \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_2$ based on $[y_j^* \mid x_j]; j = 1, 2$

$\hat{x}_2 = \frac{y^o - \hat{\theta}_{2,0}}{\hat{\theta}_{2,1}}$

Iterate $i = 2, 3, \ldots$

$x_{i+1} = x_i + \frac{1}{i^\beta \hat{\theta}_{i,1}}(y^o - y_i^*)$

$[y_{i+1}^* \mid x_{i+1}] \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_{i+1}$ based on $[y_j^* \mid x_j]; j = 1, 2, \ldots, i+1$

$\hat{x}_{i+1} = \frac{y^o - \hat{\theta}_{i+1,0}}{\hat{\theta}_{i+1,1}}$

where $\theta = (\theta_0, \theta_1)^T$ are the linear regression coefficients in the approximation $y \approx \theta^T \dot{x}$ with $\dot{x} = (1, x)^T$, $\hat{\theta}$ is the corresponding least squares estimate, and $0.5 < \beta \leq 1.0$ is a parameter of the algorithm. This algorithm is defined in Frees and Ruppert (1990), and it is demonstrated that the parametric stochastic approximation algorithm provides asymptotic pdf for the estimator:

$$\sqrt{i}\,(\hat{x}_i - x^o) \quad \underrightarrow{\mathcal{D}} \quad \mathcal{N}\left(0, \sigma_*^2/[g'(x^o)]^2\right)$$

for any sampling sequence $\{x_i\}$ converging to x^o and spread sequence $\{d_i^2\}$ diverging, which are ensured by $0.5 < \beta \leq 1.0$. Note in particular that the asymptotic pdf is independent of β.

The parametric algorithm decouples the sampling sequence $\{x_i\}$ and the estimator sequence $\{\hat{x}_i\}$. This decoupling makes it possible to have the same asymptotic estimation variance for x^o as for the optimal classical algorithms without constraints on the sampling spread sequence $\{d_i^2\}$. This will obviously be favorable for estimation of the slope $g'(x^o)$.

Consider the experimental Bayesian inversion model as defined in Section 2. There are two extensions beyond the stochastic approximation model defined above. Firstly, y^o is of lower

dimension than x and secondly, there is an error term in the likelihood model. The stochastic approximation approach can not be used on the experimental Bayesian inversion problem without fundamental revisions.

A simplified version of the experimental Bayesian inversion algorithm is very similar to the parametric stochastic approximation algorithm, however. Asymptotic results on the latter may provide guidelines for choosing the algorithm parameter sequence $\{a_i\}$ of the former. The updating procedure in the experimental Bayesian inversion algorithm is:

$$\begin{aligned} x_{i+1}^{EI} &= a_i(i+1)\hat{x}_i + [1 - a_i(i+1)]\bar{x}_i^{EI} \\ &= \bar{x}_i^{EI} + a_i(i+1)\frac{y^o - \bar{y}_i^*}{\hat{\theta}_{i,1}} \end{aligned}$$

where the relations $\hat{x}_i = \left(y^o - \hat{\theta}_{i,0}\right)/\hat{\theta}_{i,1}$ and $\hat{\theta}_{i,0} = \bar{y}_i^* - \hat{\theta}_{i,1}\bar{x}_i^{EI}$, with $\bar{y}_i^* = \frac{1}{i}\sum_{j=1}^{i} y_j^*$, are used in the development of the expression.

The corresponding updating in the parametric stochastic approximation algorithm is:

$$\begin{aligned} x_{i+1}^{SA} &= x_i^{SA} + \frac{1}{i^\beta \hat{\theta}_{i,1}}(y^o - y_i^*) \\ &= \bar{x}_i^{SA} + \frac{1}{i^\beta}\sum_{j=1}^{i}\frac{y^o - y_i^*}{\hat{\theta}_{i,1}} \end{aligned}$$

Assume that i is large enough such that $\hat{\theta}_{i,1} \approx \theta_1$, ie. the slope estimate has stablized, then:

$$\begin{aligned} x_{i+1}^{EI} &= \bar{x}_i^{EI} + a_i(i+1)\frac{y^o - \bar{y}_i^*}{\theta_1} \\ x_{i+1}^{SA} &= \bar{x}_i^{SA} + i^{1-\beta}\frac{y^o - \bar{y}_i^*}{\theta_1} \end{aligned}$$

Recall that for ensuring asymptotic unbiased Gaussianity in the parametric stochastic approximation algorithm, one must have $0.5 < \beta \leq 1.0$. Hence setting the algorithm parameter in the experimental Bayesian inversion algorithm like $\{a_i = i^{-\alpha}\}$ with $0.5 < \alpha \leq 1.0$ seems reasonable. All values of α will make the sampling sequence $\{x_i\}$ converge to x^o and make the spread design $\{d_i^2\}$ diverge. For large value of α, the convergence will be faster and divergence slower; while for small value of α, it will be the other way around. If the slope θ_1 is assumed known, $\alpha = 1.0$ will provide estimates with minimum asymptotic variance, but if θ_1 is unknown and must be estimated, it may be favorable with faster divergence in $\{d_i^2\}$ and hence a smaller α.

5 Empirical Study

The finite sample properties of the estimator for $[X \mid y^o]$ and the associated estimation variance are difficult to evaluate analytically. In this study, an evaluation through simulation is conducted. Two simulation studies are performed: one in a stochastic approximation setting and one in full experimental Bayesian inversion setting. In the former case, three different types of estimators are evaluated and compared, while in the latter only the experimental Bayesian inversion estimator is involved since this estimator is the only one applicable in this case.

5.1 Stochastic approximation study

The study consists of two parts based on different models for $g(x)$. In the first part $g(x)$ is linear and in the second it is smoothly curved. In each part, three different algorithms are evaluated and compared.

5.1.1 Algorithms

The three algorithms to be compared are similar in many ways. All algorithms are based on sequential sampling and the inversion relation is defined to be linear approximation to $g(x)$. They all start with same two initial design points that define lower and upper bounds of the design interval. The major difference is on the sequence of samples. For all algorithms a bootstrap estimate of the estimation variance is compared, which will define the stop criterion in practical applications.

D - algorithm

In this algorithm the sampling sequence corresponds to the sample design which provide D-optimality in estimating the parameters in the linear approximation to $g(x)$, Pukelsheim (1993). The algorithm is defined as:

Algorithm: D-algorithm

Initiate

$(x_1, x_2) \rightsquigarrow (x_l, x_h)$ in random order

$[y_j^* \mid x_j]; j = 1, 2 \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_2$

$\hat{x}_2 = \frac{y^o - \hat{\theta}_{2,0}}{\hat{\theta}_{2,1}}$

Iterate $i = 2, 3, \ldots$

$$x_{i+1} = \begin{cases} \mathrm{x}_l & \text{if} \quad \mathrm{x}_i = x_h \\ \mathrm{x}_h & \text{if} \quad \mathrm{x}_i = x_l \end{cases}$$

$[y^*_{i+1} \mid x_{i+1}] \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_{i+1}$

$\hat{x}_{i+1} = \frac{y^o - \hat{\theta}_{i+1,0}}{\hat{\theta}_{i+1,1}}$

Bootstrap: $\hat{\sigma}^{2X}_{i+1}$

In the algorithm, the sequential sampling alternates x_l and x_h independent of the function $g(x)$.

SA - algorithm

This algorithm is based on the parametric stochastic approximation algorithm defined in the previous section, see Frees and Ruppert (1990). The sampling sequence focuses fast on the true value of x^o and the estimates are defined through a linear approximation to $g(x)$. The algorithm is defined as:

Algorithm: SA-algorithm

Initiate

$(x_1, x_2) \rightsquigarrow (x_l, x_h)$ in random order

$[y^*_j \mid x_j]; j = 1, 2 \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_2$

$\hat{x}_2 = \frac{y^o - \hat{\theta}_{2,0}}{\hat{\theta}_{2,1}}$

Iterate $i = 2, 3, \ldots$

$x^p = x_i + \frac{y^o - y^*_i}{i\hat{\theta}_{i,1}}$

$$x_{i+1} = \begin{cases} \mathrm{x}_l & \text{if} \quad \mathrm{x}^p < x_l \\ \mathrm{x}^p & \text{if} \quad \mathrm{x}_l \leq x^p \leq x_h \\ \mathrm{x}_h & \text{if} \quad \mathrm{x}^p > x_l \end{cases}$$

$[y^*_{i+1} \mid x_{i+1}] \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_{i+1}$

$\hat{x}_{i+1} = \frac{y^o - \hat{\theta}_{i+1,0}}{\hat{\theta}_{i+1,1}}$

Bootstrap: $\hat{\sigma}^{2X}_{i+1}$

Hence the sampling sequence is according to the adaptive Robbins-Monro procedure with $\beta = 1.0$, see Lai and Robbins (1981), although forced to be in the interval $[x_l, x_h]$.

EI - algorithm

This is the reduced version of the experimental Bayesian inversion algorithm suited for the stochastic approximation study. The algorithm is defined as follows:

Algorithm: EI-algorithm

Initiate

Set algorithm parameter $0.5 < \alpha \leq 1.0$

$(x_1, x_2) \rightsquigarrow (x_l, x_h)$ in random order

$\bar{x}_{(2)} = \frac{1}{2}(x_1 + x_2)$

$[y^*_j \mid x_j]; j = 1, 2 \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_2$

$\hat{x}_2 = \frac{y^o - \hat{\theta}_{2,0}}{\hat{\theta}_{2,1}}$

Iterate $i = 2, 3, \ldots$

$x^p = \frac{1}{i^\alpha}(i+1)\hat{x}_i + [1 - \frac{1}{i^\alpha}(i+1)]\bar{x}_i$

$$x_{i+1} = \begin{cases} \mathrm{x}_l & \text{if } \mathrm{x}^p < x_l \\ \mathrm{x}^p & \text{if } \mathrm{x}_l \leq x^p \leq x_h \\ \mathrm{x}_h & \text{if } \mathrm{x}^p > x_l \end{cases}$$

$\bar{x}_{i+1} = \frac{1}{i+1}\{i\bar{x}_i + x_{i+1}\}$

$[y^*_{i+1} \mid x_{i+1}] \rightsquigarrow f(y^* \mid x)$

Compute $\hat{\theta}_{i+1}$

$\hat{x}_{i+1} = \frac{y^o - \hat{\theta}_{i+1,0}}{\hat{\theta}_{i+1,1}}$

Bootstrap: $\hat{\sigma}^{2X}_{i+1}$

The sequential sampling is forced to be within the interval $[x_l, x_h]$. The algorithm is evaluated for three different values of α with $0.5 < \alpha \leq 1.0$.

5.1.2 Study design

There are two parts of the study design defined by the shape of $g(x)$, one linear and one smoothly curved. In each part, the size and geometry of the window $[x_l, x_h]$ relative to the true value x^o and the variance of the experiment σ_*^2 are varied.

Part I. Linear model

This part is defined by the following model:

$$x, y \in \mathcal{R}^1$$
$$g(x) = x$$
$$x^o = 10,\ y^o = 10,\ \sigma_o^2 = 0$$
$$f(x) \rightarrow Uni[x_l, x_h]$$
$$f(y^* \mid x) \rightarrow \mathcal{N}(g(x), \sigma_*^2)$$

The study design in this part is:

Design no.	I	II	III	IV	V	VI
x_l	5	-40	5	5	-40	5
x_h	15	60	60	15	60	60
σ_*^2	1.0	1.0	1.0	5.0	5.0	5.0

Note that the uniform prior model on x can be interpreted as the range of x for which the linear model is assumed to be valid.

This linear case is, of course, very particular since the inversion in all algorithms is based on a linear approximation to $g(x)$. In this case the approximation is exact. The results for this part is therefore expected to be better than those for general non-linear cases.

In the current study $\theta = (0, 1)$ since $g(x) = x$. All algorithms provide $\hat{\theta}_i$ for arbitrary i that are unbiased for θ. The D-algorithm provide minimum variance estimates for θ. All estimators $\hat{\theta}_i$ are consistent estimators for θ when $i \rightarrow \infty$. The estimator for x^o, $\hat{x}_i$, at arbitrary i will be biased due to non-linear dependence on θ. The bias and variance of this estimator as function of i, $(\hat{x}_i - x^o)$ and σ_i^{2X}, is the target of this study. It is known, however, that $\hat{x}_i$ is a consistent estimator for x^o when $i \rightarrow \infty$ for all algorithms, since $\hat{x}_i$ is a smooth function of $\hat{\theta}_i$ which is consistent. Hence all algorithms are assumed to perform well asymptotically in this case. The finite sample properties are more difficult to assess, and these properties are the focus of the study.

Part II. Curved model

This part is defined by the following model:

$$x, y \in \mathcal{R}^1$$

$$g(x) = x + 0.05\delta(x - x^o) \text{ with } \delta(t) = \begin{cases} 0.1\text{t}^2 & \text{if t} \geq 0 \\ \text{t}^2 & \text{else} \end{cases}$$

$$x^o = 10,\ y^o = 10,\ \sigma_o^2 = 0$$

$$f(x) \rightarrow Uni[x_l, x_h]$$

$$f(y^* \mid x) \rightarrow \mathcal{N}(g(x), \sigma_*^2)$$

The study design in this part is:

Design no.	I	III	IV	VI
x_l	5	5	5	5
x_h	15	60	15	60
σ_*^2	1.0	1.0	5.0	5.0

The shape of $g(x)$ is displayed in Figure 1. Note that the uniform prior model on x can be interpreted as the range of x for which one can assume that $\frac{\partial}{\partial x} g(x) > 0; \forall x \in (x_l, x_h)$. This property is a general requirement for applying stochastic approximation algorithms. Under the current model for $g(x)$, few characteristics for the estimators produced from the algorithms are analytically available. It is know, however, that the estimators $\hat{x}_i$ for x^o in the SA-algorithm and the EI-algorithm are consistent when $i \rightarrow \infty$, see Frees and Ruppert (1990) and the discussion in the previous section. The objective is to evaluate the bias $(\hat{x}_i - x^o)$ and estimation variance σ_i^{2X} in the finite sample case, however.

5.1.3 Empirical procedure

The simulation study is conducted as follows:

1. For a given model, study design and algorithm, 300 simulations are run resulting in:

   - 300 sampling sequences $\{x_i\}$. Example of one sequence is presented in the upper-left display in Figure 2.
   - 300 sampling sequences $\{(x_i, y_i^*)\}$. Example of one sequence is presented in the upper-right display in Figure 2.
   - 300 estimate sequences $\{\hat{x}_i\}$. Example of these sequences is presented in the second-left display in Figure 2.
   - 300 estimated squared error sequences $\{(\hat{x}_i - x^o)^2\}$. Example of these sequences is presented in the third-left display in Figure 2.

- 300 bootstrap estimates of estimation variance $\{\hat{\sigma}_i^{2X}\}$, based on 200 bootstrap replications. Example of these sequences is presented in the lower-left display in Figure 2.

2. Average of the generated sequences over these 300 simulations are calculated. Examples of these are presented in the right-column from second and down in Figure 2.

5.1.4 Results and discussion

For a given model and study design, the results provided by the three algorithms are summarized in one figure, see Figure 3 as an example. Each column in the figure corresponds to one algorithm. The two upper displays represent one simulation of sampling sequence $\{x_i\}$ and the elements in $\{(x_i, y_i^*)\}$, respectively. The third through fifth display in Figure 3 represent average over the 300 samples of the estimates, estimated squared error and bootstrap estimate of estimation variance, respectively. The lower display is a cross-plot of the bootstrap estimate of estimation variance versus the average squared error (ie, as vertical axis versus horizontal axis).

Results: Part I. Linear model

Figure 3 through 8 present the results for the D-algorithm, SA-algorithm and EI-algorithm with $\alpha = 1.0$. The latter is used because it is asymptotically comparable with the SA-algorithm. Figure 9 through 11 present the results for the EI-algorithm with varying α, ie $\alpha = 0.51, 0.75, 1.0$. Figure 12 presents a summary of the figures above.

Figure 3 displays the results from Design I, which is a narrow symmetric domain with small experimental variance. Note that the sampling sequences of the D-algorithm remain on the border of the domain while those of the two other algorithms focus on $x^o = 10$. The estimate and the square error sequences for all the three algorithms appear as comparable and very reliable. Moreover, the bootstrap estimates of estimation variance appear as unbiased for squared error for all algorithms.

Figure 4 displays the results from Design II, which is a wide symmetric domain with small experimental variance. The results are almost identical to the ones obtained from Design I. Hence all algorithms perform similarly.

Figure 5 displays the results from Design III, which is a wide asymmetric domain with small experimental variance. The sampling sequences are as expected: the D-algorithm remain on the border of the domain while that of SA-algorithm and the EI-algorithm focus fast on x^o. The empirical squared error sequence shows that the D- and SA-algorithm are very similar. The EI-algorithm appears significantly better than the two others. The bootstrap estimates appear reliable for all the three algorithms.

The D-algorithm is optimal for estimating $\theta = (0, 1)$, but predicting y^o at x^o is a different problem. The prediction variance is dependent on the deviation between x^o and the centre of the sampling sequence. The centre of the D-algorithm sampling sequence is far off x^o. It is at

27.5. This lack of centering around x^o makes the D-algorithm inefficient in solving the inverse problem. The SA-algorithm produces a sampling sequence that eventually focuses at x^o, but the average of the path is not well centred at x^o. The sampling sequence of the EI-algorithm focuses in at x^o, and its average is also well centred at x^o. This centering is expected to be crucial such that the EI-estimates appear as superior to the others.

Figure 6 through 8 display the results from Design IV, V and VI, which correspond to Design I, II and III except that a large experimental variance is considered. Largely, the same conclusions can be drawn from these figures as from Figure 3 through 5. There appears to be a lack of stability in the bootstrap estimates for the SA-algorithm and EI-algorithm when the ratio of experimental variance over domain width is large, and this unstability is probably caused by the poor sampling design in estimating θ.

Figure 9 through 11 display the results from evaluation of the EI-algorithm with varying value of the algorithm parameter α. Recall that $0.5 < \alpha \leq 1.0$ in order to ensure consistency in the estimates for non-linear $g(x)$, which entails that all sampling sequences focus in at x^o. Note, however, that this convergence is much slower for small values of α. For linear $g(x)$, this slow convergence seems to improve the estimate slightly, probably because the slope can be more reliably estimated. Moreover, small values of α will provide sampling paths where the bootstrap estimates of estimation variance are much more stable.

Figure 12 summarizes the results from the empirical study under the linear model. In practice one will expect that the domain is relatively wide and the experimental variance is small. Moreover, the solution will normally fall asymmetrically in the domain. The most realistic design will hence be Design III. For this case, the EI-algorithm is clearly favorable to the D- and SA-algorithm regardless of the choice of $0.5 < \alpha \leq 1.0$. A small value of α seems to provide slightly better estimates than a large one, however. Moreover, bootstrap estimates of estimation variance appear as more reliable for small α.

The empirical study on Part I. Linear model can be summarized as follows. The EI-algorithm regardless of value of $0.5 < \alpha \leq 1.0$ seems to be uniformly dominant over the two other algorithms, ie it has smaller or equal empirical squared error sequence for all cases. Small value of α may be slightly preferable. All algorithms performed reasonably well for domains that are symmetric around x^o. If x^o falls asymmetrically in the domain, the EI-algorithm is clearly preferable. Bootstrap estimates of expected square error appear as very reliable and may be used as stop criterion in the inversion. One exception is for narrow domain with high experimental variance, where the bootstrap estimates seem to be unstable.

Results: Part II. Curved model

Figure 13 through 16 present the results for the D-algorithm, SA-algorithm and EI-algorithm with $\alpha = 1.0$. The latter is used because it is asymptotically comparable with the SA-algorithm. Figure 17 presents the results for the EI-algorithm with varying α. Figure 18 presents a summary of the figures above.

Figure 13 displays the results from Design I, which is a narrow symmetric domain with small experimental variance. The sampling sequences are similar to the ones for the linear model.

The results are, however, very different since the skewness in the estimates of x^o appears as critical. The D-algorithm provides dramatically skewed estimates of the x^o since θ is estimated based on samples at the borders of the domain only. Note further that the skewness is not corrected as $i \to \infty$. The two other algorithms focus the sampling sequence to x^o fairly fast, hence the skewness is corrected fairly fast. Both the SA- and EI-algorithm provide reliable estimates and the bootstrap estimates of estimation variance are largely representative.

Figure 14 displays the results from Design III, which is a wide asymmetric domain with small experimental variance. Here the convergence problems are larger since the non-linearity is more pronounced. The D-algorithm is completely unreliable. Both SA- and EI-algorithm appear to provide estimates which converge towards x^o according to theory. The convergence of the SA-algorithm is faster than that for the EI-algorithm, but for both algorithms the bootstrap estimates of estimation variance are downward biased.

Figure 15 and 16 correspond to Figure 13 and 14, respectively, except for the experimental variance which is increased. The picture is largely the same, except for lack of stability in bootstrap estimates for the narrow domains with large experimental variance.

Figure 17 displays results from the evaluation of the EI-algorithm with varying value of α on Design III. The results are very comparable for $\alpha = 0.51, 0.75, 1.0$, although $\alpha = 1.0$ appear as slightly best.

Figure 18 summarizes the results from the empirical study under the non-linear model. In practice, the non-linearity may be pronounced and the experimental variance may relatively be small. Moreover, the solution will normally fall asymmetrically in the domain. The most realistic design will hence be Design III. The D-algorithm is completely useless. The SA- and EI-algorithm, for all values of $0.5 < \alpha \leq 1.0$, converge correctly. The SA-algorithm appears with fastest convergence while for the EI-algorithm large α-value should be used. Moreover, bootstrap estimates appear as severely downward biased.

5.2 Experimental Bayesian inversion study

The study consists of one model with $g(x)$ being linear and different designs. Only the experimental Bayesian inversion algorithm is evaluated since it is the only algorithm applicable for this case.

5.2.1 Algorithm

The experimental Bayesian inversion algorithm is defined in Section 3. The version used here is based on algorithm parameters $\{a_i = i^{-1}\}$, ie $\alpha = 1.0$. This ensures convergence even for non-linear $g(x)$ in the stochastic approximation setting and is expected to do so in the experimental Bayesian inversion. In addition, an algorithm with sampling sequence generated independently from the prior pdf of x is evaluated. It is termed prior-sampling experimental Bayesian inversion algorithm. This corresponds to the naive choice of sampling sequence which focuses on estimating the parameters in $g(x)$.

5.2.2 Study design

The model is as follows:

$$
\begin{aligned}
&x \in \mathcal{R}^2,\ y \in \mathcal{R}^1 \\
&g(x) = a^T x \text{ with } a = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \\
&y^o = 20,\ \sigma_o^2 = 1.0 \\
&f(x) \to \mathcal{N}_2(\mu_X, \Sigma_X) \text{ with } \mu_X = \begin{bmatrix} \mu_{X_1} \\ \mu_{X_2} \end{bmatrix} \text{ and } \Sigma_X = \begin{bmatrix} \sigma^2_{X_1} & 0 \\ 0 & \sigma^2_{X_2} \end{bmatrix} \\
&f(y^* \mid x) \to \mathcal{N}_1(g(x), 9.0)
\end{aligned}
$$

and the study designs are:

Design no.	I	II	III	IV	V
μ_{X_1}	10	-5	5	4	4
μ_{X_2}	10	25	5	6	6
$\sigma^2_{X_1}$	9	9	9	9	9
$\sigma^2_{X_2}$	9	9	9	9	4

5.2.3 Empirical procedure

The empirical procedure is largely the same as in Section 5.1.3, but since the dimension of x is larger the displays in the figures are somewhat different.

5.2.4 Results and discussion

For a given study design, the results provided by the experimental Bayesian inversion algorithm are summarized in one figure, see Figure 19 as an example. Top left presents the elements in the sampling sequence $\{x_i\}$ in the $x = (x_1, x_2)$ - plane, while top right displays the experimental value sequence $\{y_i^*\}$. Row two and three display the results for the x_1-dimension. Row two contains the estimate sequence $\{\hat{x}_{1i}\}$ for the 300 realizations and the average over them. Row three contains the bootstrap estimate of estimation variance for the 300 realizations and the average over them. Row four and five illustrate the corresponding plots for the x_2-dimension. The last row displays the bootstrap estimate of the estimation variance including uncertainty in $g(x)$ divided by estimation variance given $g(x)$, which corresponds to the stop criterion. This ratio will, of course, approach unity as $i \to \infty$ since the parameters in $g(x)$ will be consistently estimated.

Figure 19 contains the results from Design I, which has a prior model that is correctly centred. The samples are centred around the correct value $(10, 10)$ although having some variability. This is not surprising as the prior is centred at the plausible solution. The estimator sequence

$\{\hat{x}_i\}$ converges very fast to the correct value $(x_1, x_2) = (10, 10)$. The bootstrap estimated variance is close to the theoretically correct value.

Figure 20 contains the results from Design II. Again the centre of the prior $(-5, 25)$ is a plausible solution since it adds up to 20. The results are very similar to the ones in Figure 19, ie fast convergence to the correct solution both for estimate and estimation variance.

Figure 21 contains the results from Design III which has a prior expectation that can not be a solution. The prior model is symmetric in (x_1, x_2), however, and one observes that the estimate is somewhat below 10 for both x_1 and x_2. The observation error in y^o provides freedom for this. The estimation variance is consistently determined by the bootstrap estimate, however, although a number of samples are required to ensure convergence.

Figure 22 contains the results from Design IV which has asymmetric prior. Hence the estimates for x_1 and x_2 are different, one above and the other below 10. Again estimation variance appear to be consistently determined by the bootstrap estimate.

Figure 23 contains the results from Design V where both expectation and variance vary in the (x_1, x_2)-dimensions of the prior. The results correspond to the ones in Figure 22, although convergence of the bootstrap estimates are somewhat slower.

Figure 24 contains the results from the prior-sampling experimental Bayesian inversion from Design III. Recall that this is the naive approach to solving the experimental Bayesian inversion problem focusing on estimating the parameter in $g(x)$. The convergence as $i \to \infty$ is the same as in the comparable Figure 21. The convergence rate is considerably slower, however. Note how the sampling sequence in the upper left display of both figures vary. In Figure 21 it is centred around the solution $[x \mid y^o]$ while in Figure 24 it is centred around the prior expectation μ_X.

The empirical study in the experimental Bayesian inversion setting can be summarized as follows. The experimental Bayesian inversion algorithm provides consistent estimates of the posterior expectation and posterior variance and the convergence is fast. It is much more efficient than a naive algorithm that focuses on estimating the parameter of $g(x)$ instead of solving the inverse problem, see Figure 25.

6 Conclusions

Experimental inversion is defined as a traditional inversion problem with the additional complication that the forward model is not directly accessible. It can only be determined including an error term. By casting the problem in a Bayesian setting, the experimental Bayesian inversion problem is defined. Bayesian inversion is usually solved through MCMC or MAP approaches, but these approaches are not directly applicable to the extended inversion problem. Hence a new experimental Bayesian inversion algorithm is defined. A simplified version of this algorithm has several similarities with stochastic approximation algorithms as defined in literature.

An empirical simulation study is conducted in order to evaluate the finite sample properties of the estimators which are defined by the algorithms. Focus is on solving the inverse problem

at a pre-defined precision by a minimum number of calls of the forward model. The study has two parts: stochastic approximation setting and experimental Bayesian inversion setting.

The stochastic approximation setting consists of a special case of the experimental Bayesian inversion, and it includes a simplified well-posed inversion problem. Both linear and non-linear forward models are studied. The conclusions from the study are as follows. The experimental Bayesian inversion algorithm appears to be favorable to all stochastic approximation algorithms presented in literature for linear forward models. Both the experimental Bayesian inversion and stochastic approximation algorithms provide consistent estimates for non-linear forward models, but the latter appears to have a slightly faster convergence rate.

It is worth noting that the experimental Bayesian inversion algorithm is constructed to solve complex ill-posed inverse problems, and hence much more general than traditional stochastic approximation algorithms.

The experimental inversion setting includes an underdetermined inverse problem with a linear forward model. The conclusions from the study are as follows. The experimental Bayesian inversion algorithm provides consistent estimates of conditional expectation and conditional variance. The convergence rate is fast. The experimental Bayesian inversion algorithm appears to be more efficient than traditional algorithms which focus on estimating the forward model first for thereafter to perform the inversion.

7 References

Frees, E. and D. Ruppert (1990). Estimation following a sequentially designed experiment. *Journal of Amer. Statistical Association*, 85, 412.

Ghosh, M., N. Mukhopadhyay and P.K. Sen (1997). *Sequential estimation*. John Wiley & Sons, Inc., New York.

Hansen, P.C. (1998). Rank-deficient and discrete ill-posed problems. SIAM Monographs on Mathematical Modeling and Computation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA.

Kolbjørnsen, O. (2002). Nonlinear topics in the Bayesian approach to inverse problems with applications to seismic inversion. Doctoral Dissertation, Department of Mathematical Sciences, Norwegian University of Science and Technology, Trondheim.

Lai, T. and H. Robbins (1979). Adaptive designs and stochastic approximation. *The annals of Statistics*, 7, 1196-1221.

——————(1981). Consistency and asymptotic efficiency of slope estimates in stochastic approximation schemes. *Zeitschrift fr Wahrscheinlichkeitstheorie und Verwandte Gebiete*, 56, 329-366.

Mardia, K.V., J.T. Kent and J.M. Bibby (1979). *Multivariate analysis*. Academic Press Inc., London.

Pukelsheim, F. (1993). *Optimal design of experiments*. John Wiley & Sons, Inc., New York.

Robbins, H. and S. Monro (1951). A stochastic approximation method. *Annals of Mathematical statistics*, 22, 400-407.

Robert, C.P. and G. Cassela (1999). *Monte Carlo statistical methods*. Springer Texts in Statistics.

Springer-Verlag, New York.

Wei, C. (1985). Asymptotic properties of least squares estimation in stochastic regression models. *The annals of statistics*, 13, 1498-1508.

Appendix - Figures

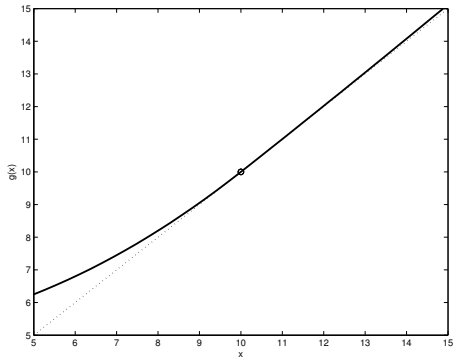


Figure 1: Plot of the curved model for $g(x)$ (solid line).

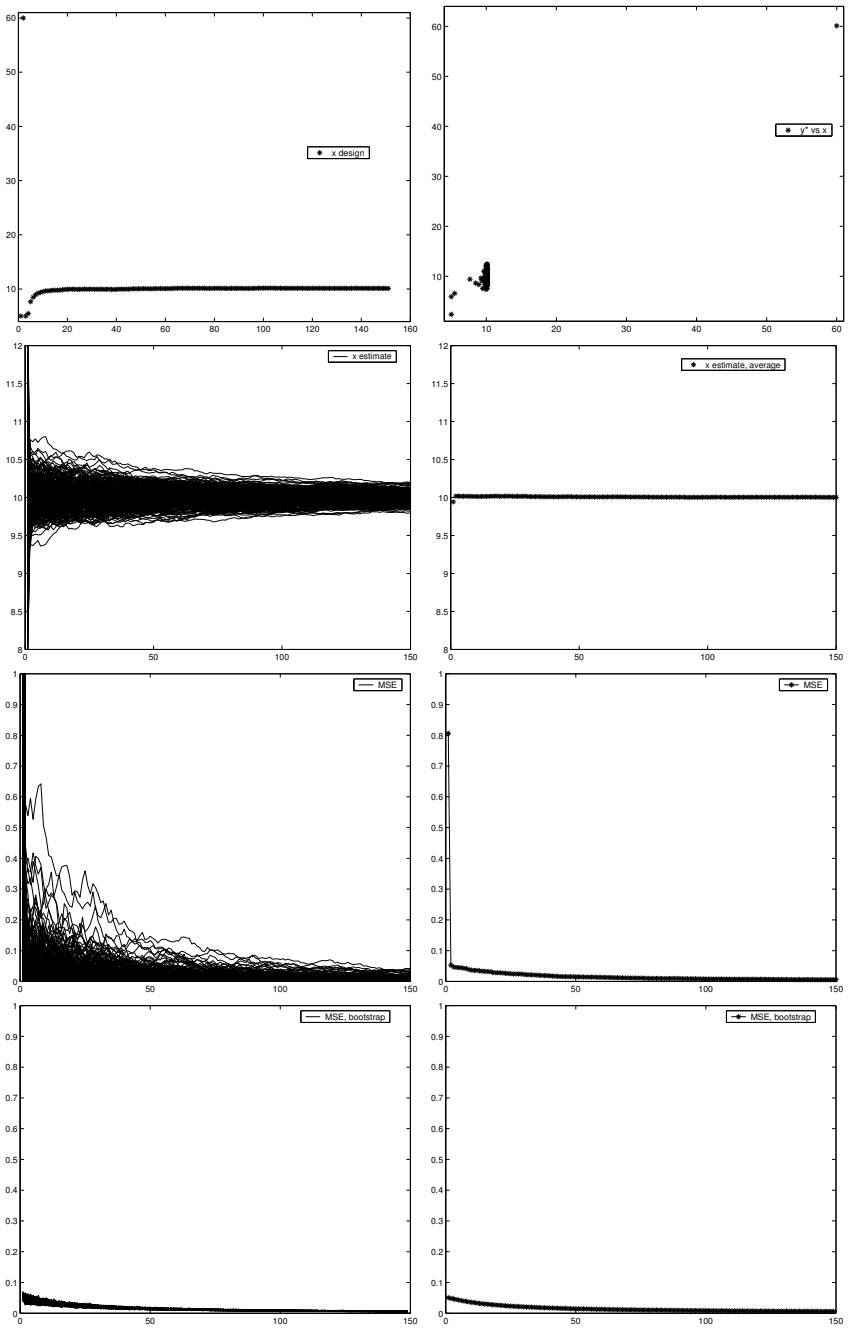


Figure 2: Simulation results of the EI-algorithm with $\alpha = 1.0$ based on the linear model /Design III, ie with $(x_l, x_h) = \{5, 60)$ and $\sigma_*^2 = 1.0$.

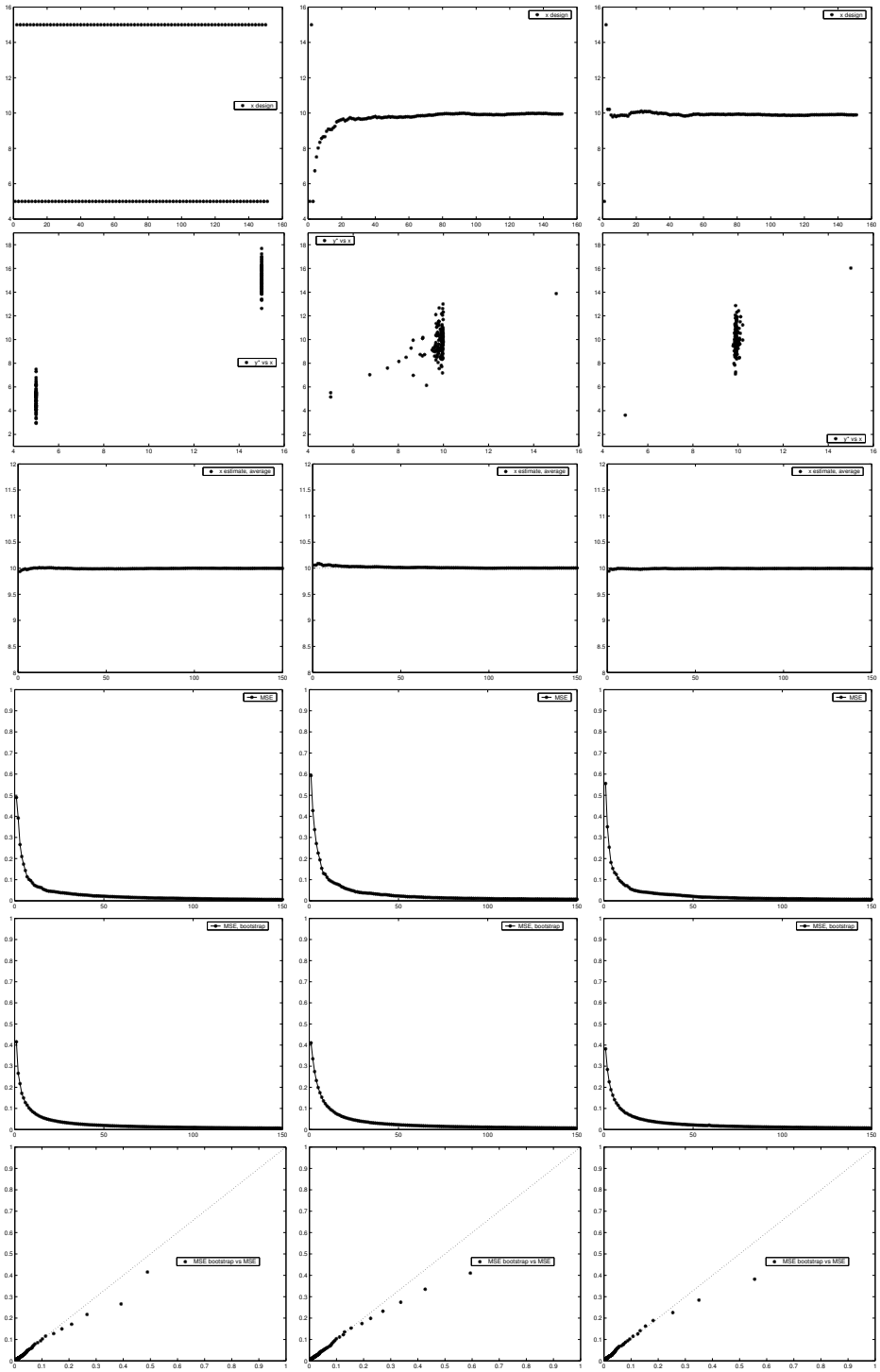


Figure 3: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the linear model /Design I, ie $(x_l, x_h) = (5, 15)$ and $\sigma_*^2 = 1.0$.

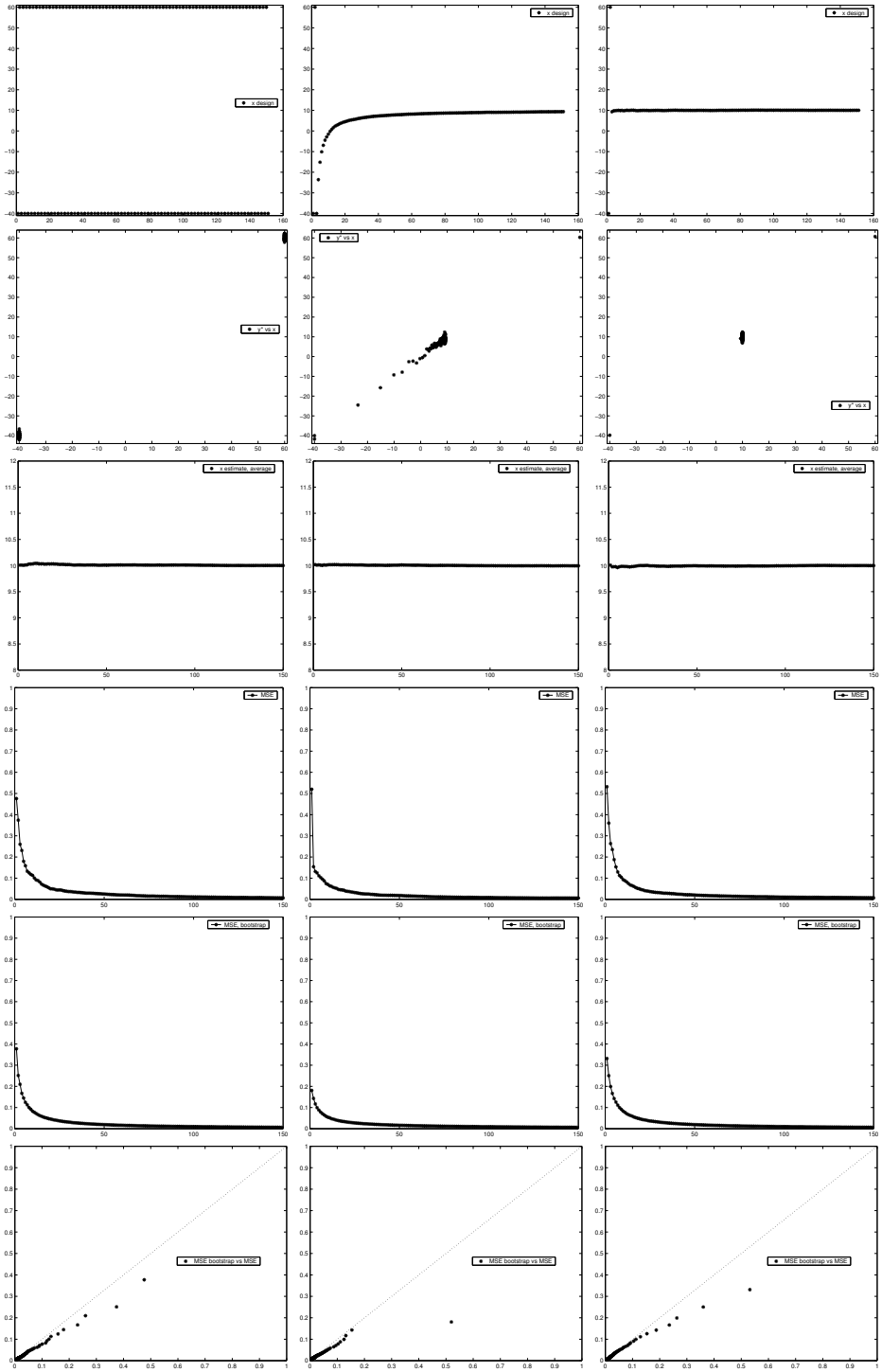


Figure 4: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the linear model /Design II, ie $(x_l, x_h) = (-40, 60)$ and $\sigma_*^2 = 1.0$.

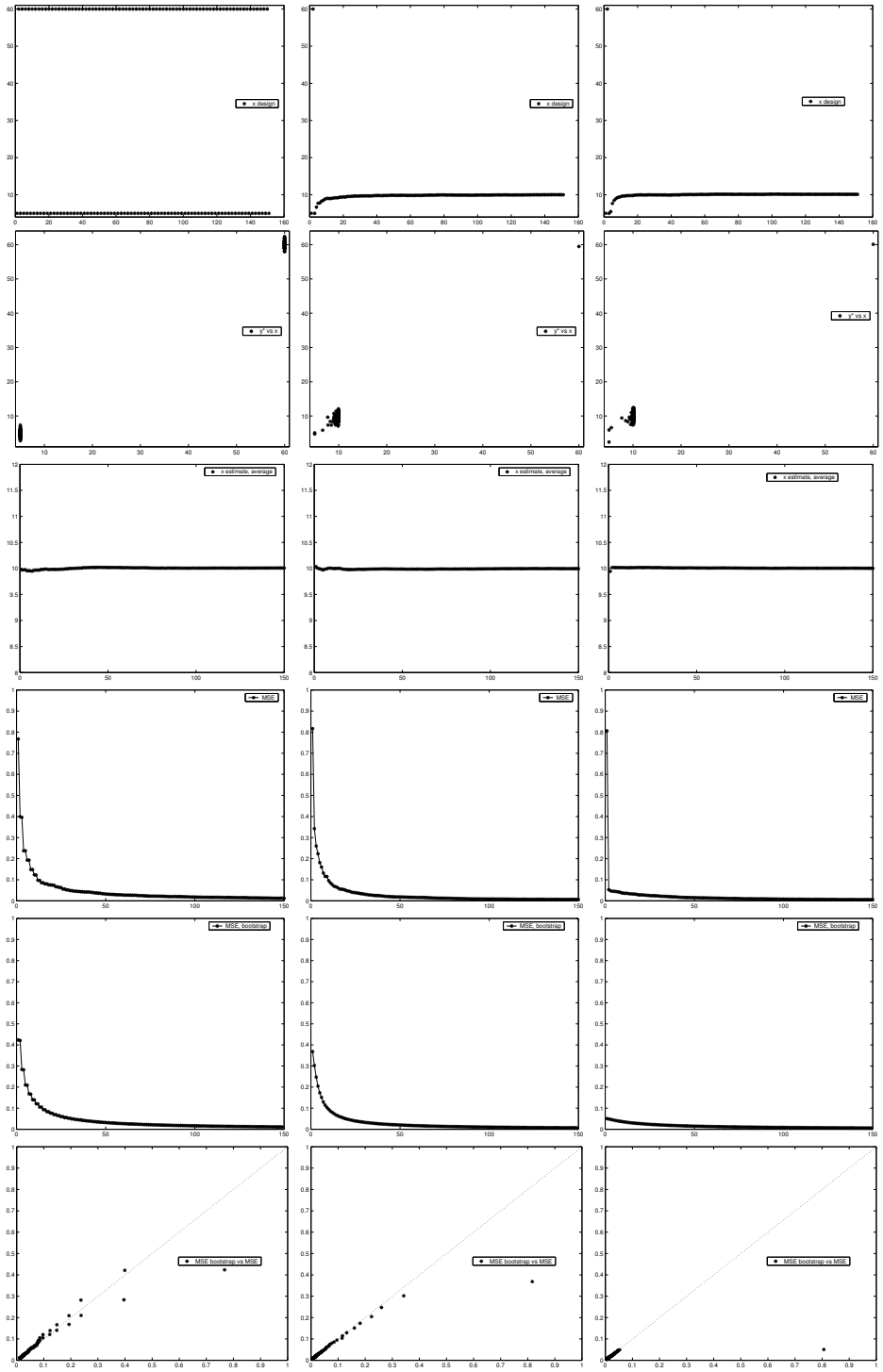


Figure 5: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the linear model /Design III, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 1.0$.

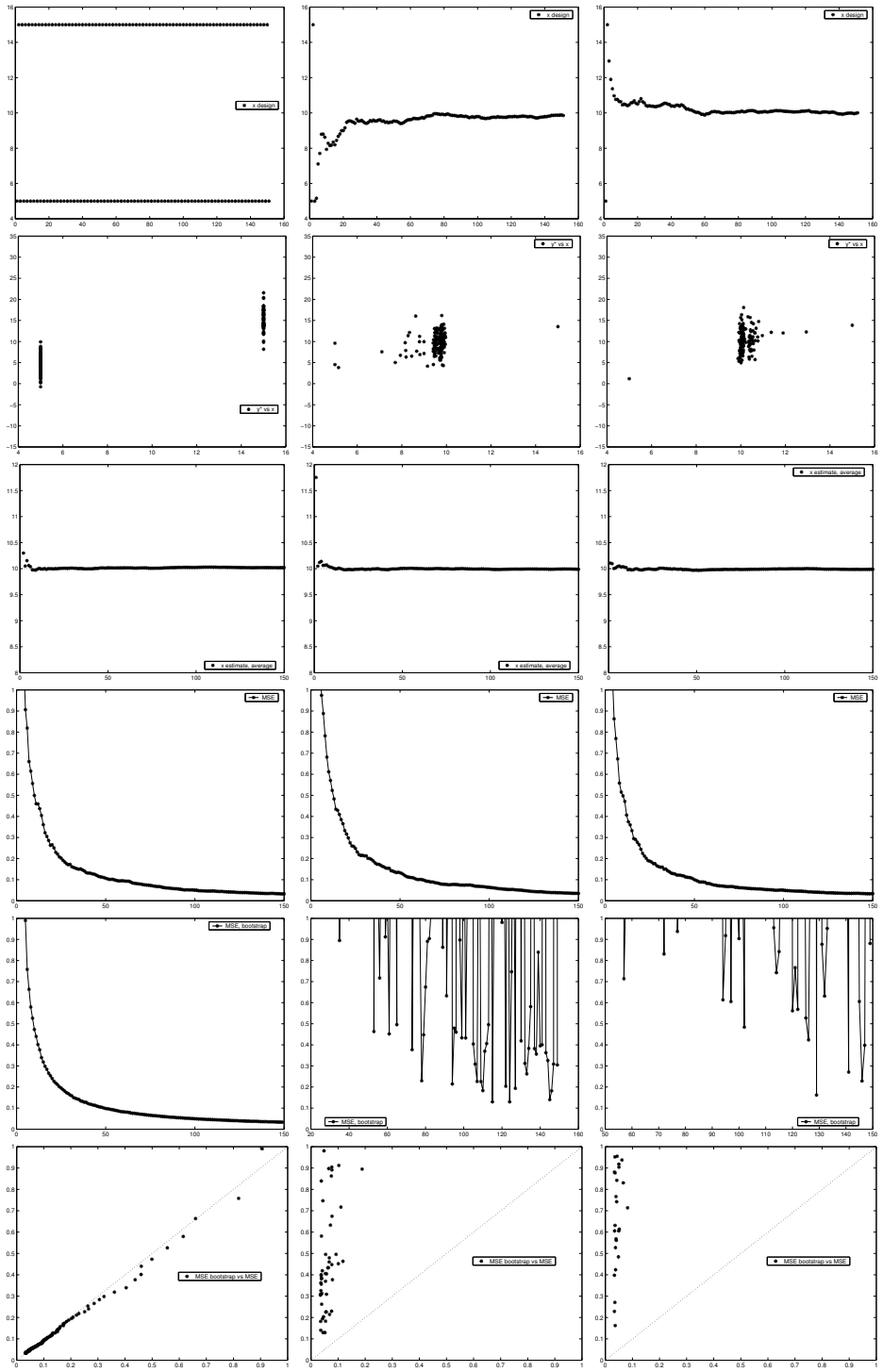


Figure 6: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the linear model/Design IV, ie $(x_l, x_h) = (5, 15)$ and $\sigma_*^2 = 5.0$.

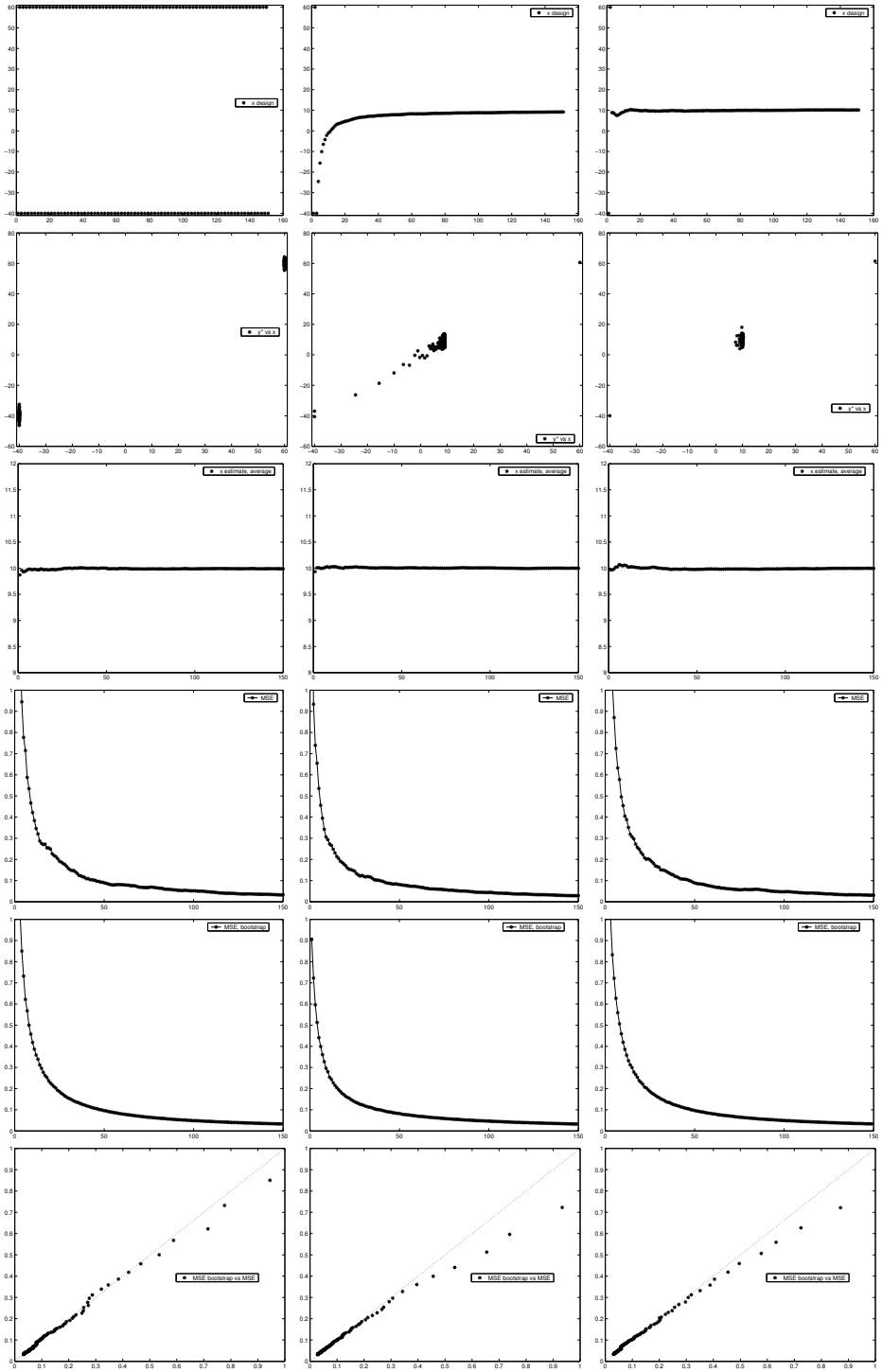


Figure 7: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the linear model/Design V, ie $(x_l, x_h) = (-40, 60)$ and $\sigma_*^2 = 5.0$.

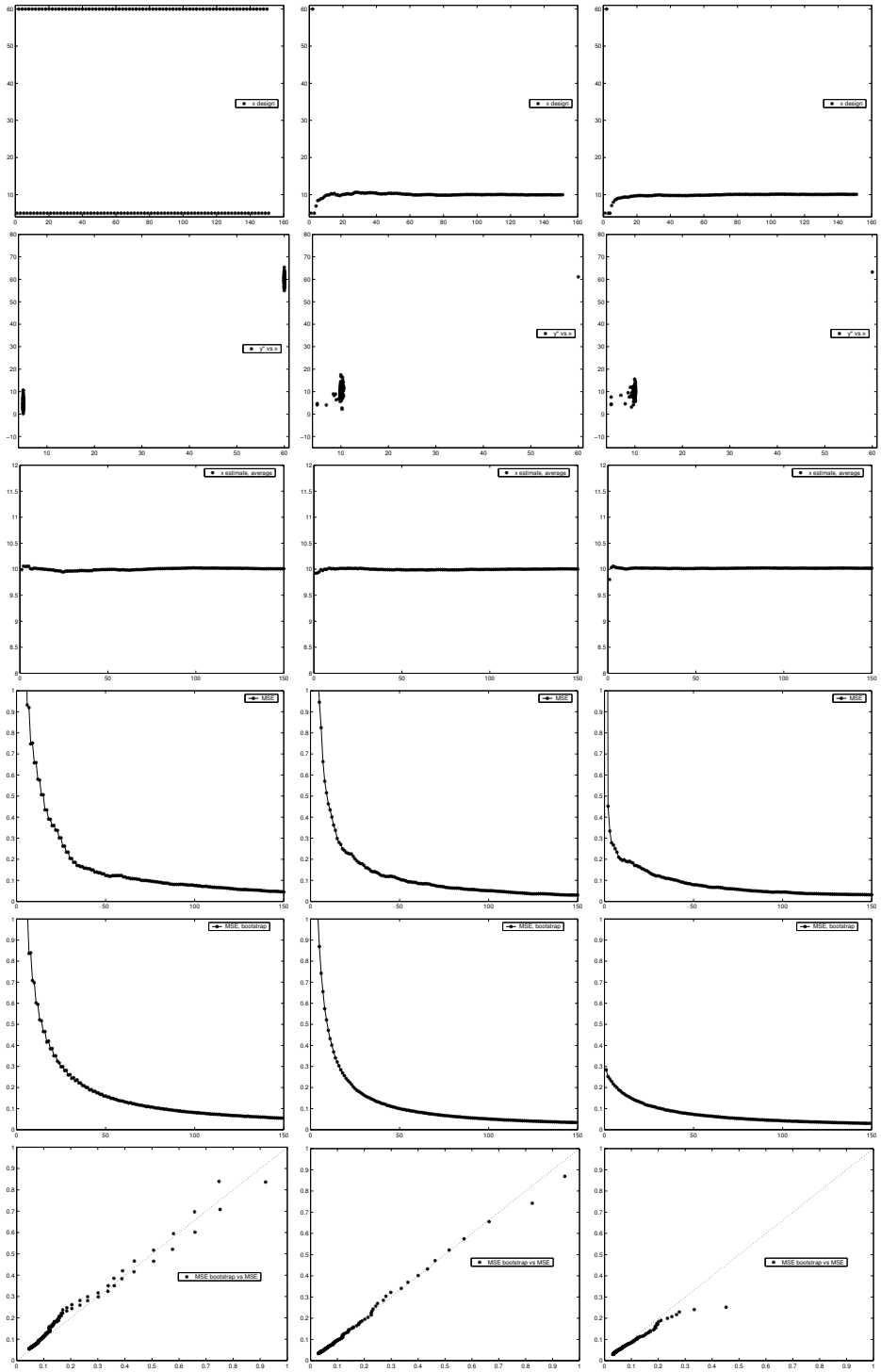


Figure 8: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the linear model/Design VI, $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 5.0$.

$\alpha = 0.51$ $\alpha = 0.75$ $\alpha = 1.0$

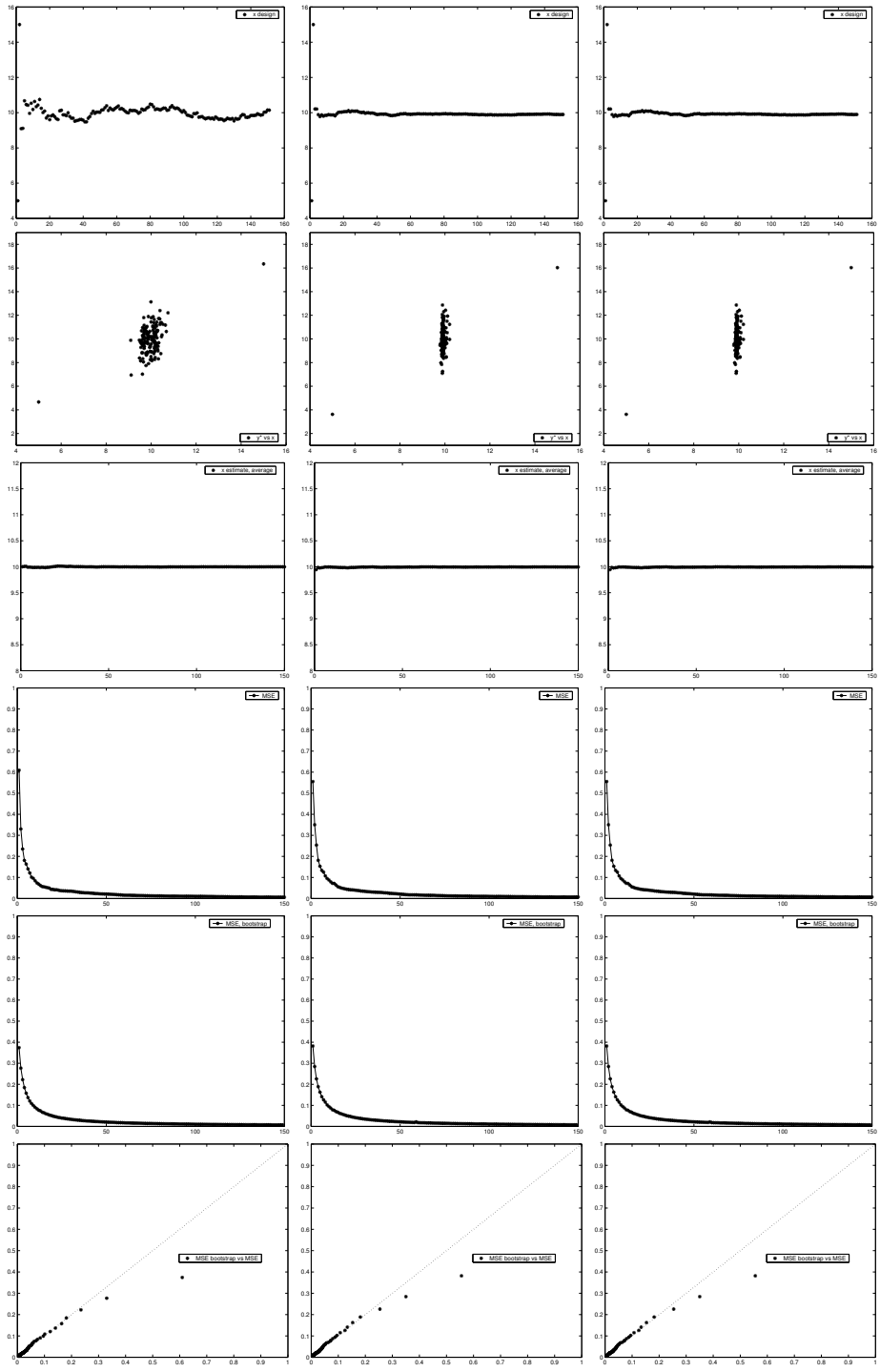


Figure 9: Simulation results of the EI-algorithm with varying α based on the linear model/Design I, ie $(x_l, x_h) = (5, 15)$ and $\sigma_*^2 = 1.0$.

$\alpha = 0.51$ $\alpha = 0.75$ $\alpha = 1.0$

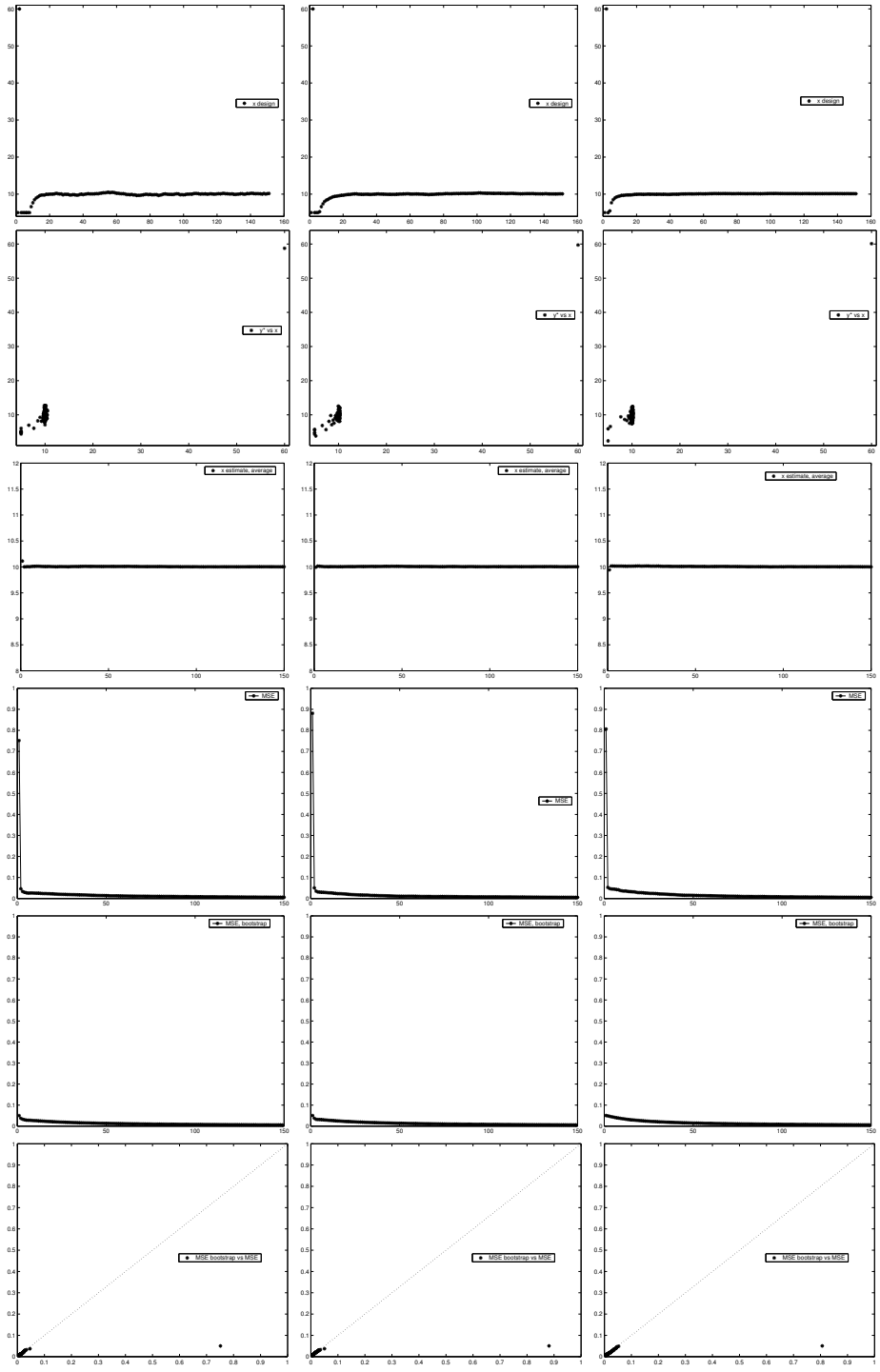


Figure 10: Simulation results of the EI-algorithm with varying α based on the linear model/Design III, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 1.0$.

$\alpha = 0.51$ $\alpha = 0.75$ $\alpha = 1.0$

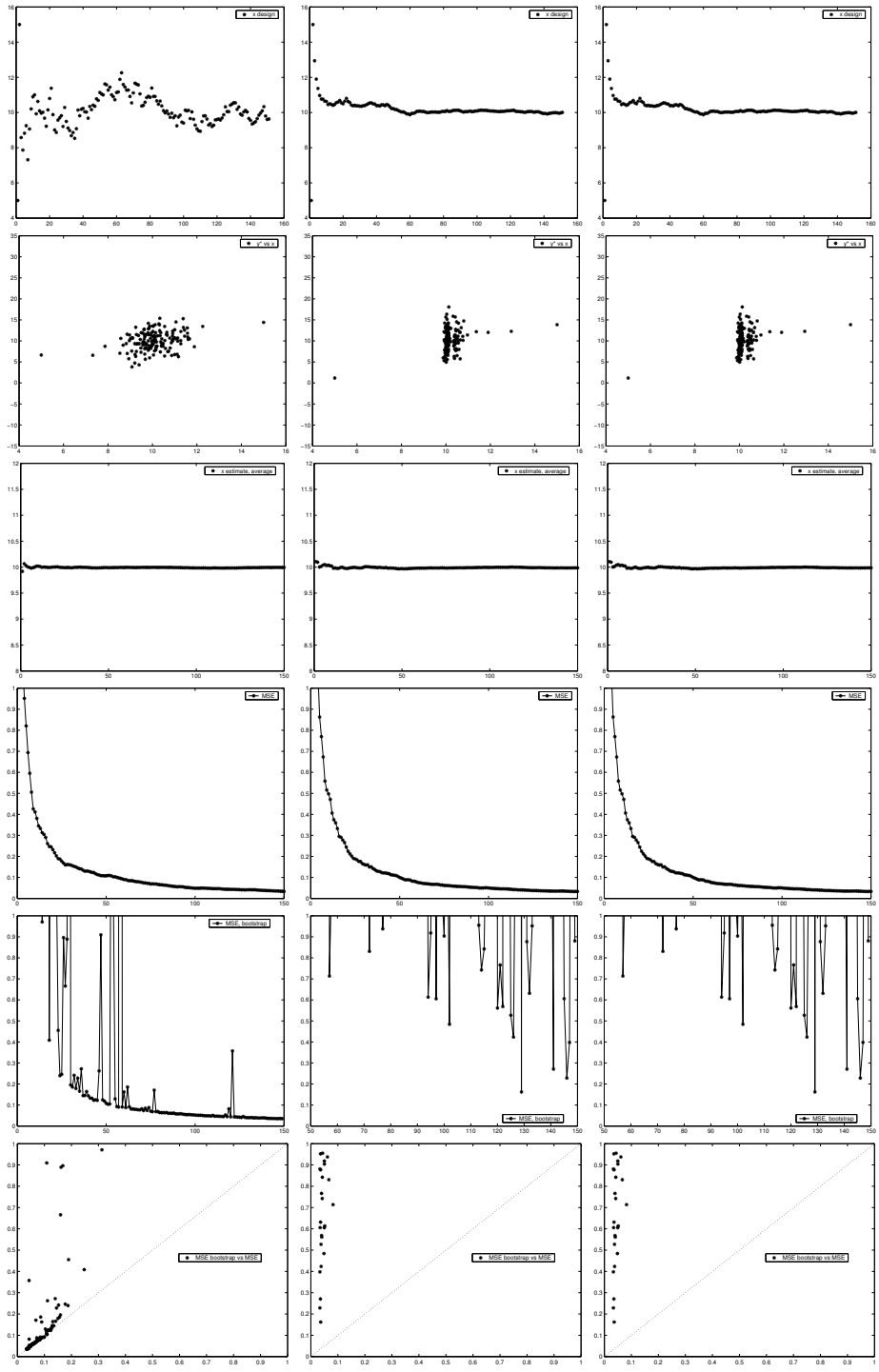


Figure 11: Simulation results of the EI-algorithm with varying α based on the linear model/Design IV, ie $(x_l, x_h) = (5, 15)$ and $\sigma_*^2 = 5.0$.

$\alpha = 0.51$ $\alpha = 0.75$ $\alpha = 1.0$

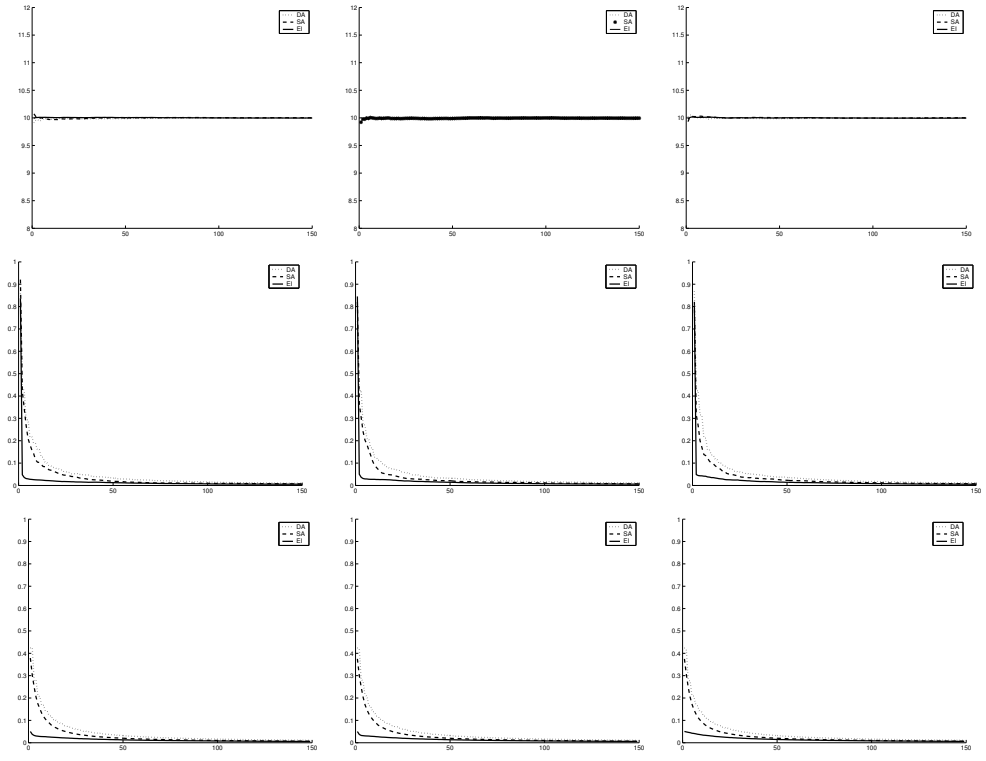


Figure 12: Comparison plots for D-, SA- and EI-algorithm with varying α based on the linear model/Design III, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 1.0$. The displays from top to bottom are estimates, estimated squared error and bootstrap estimate of estimation variance.

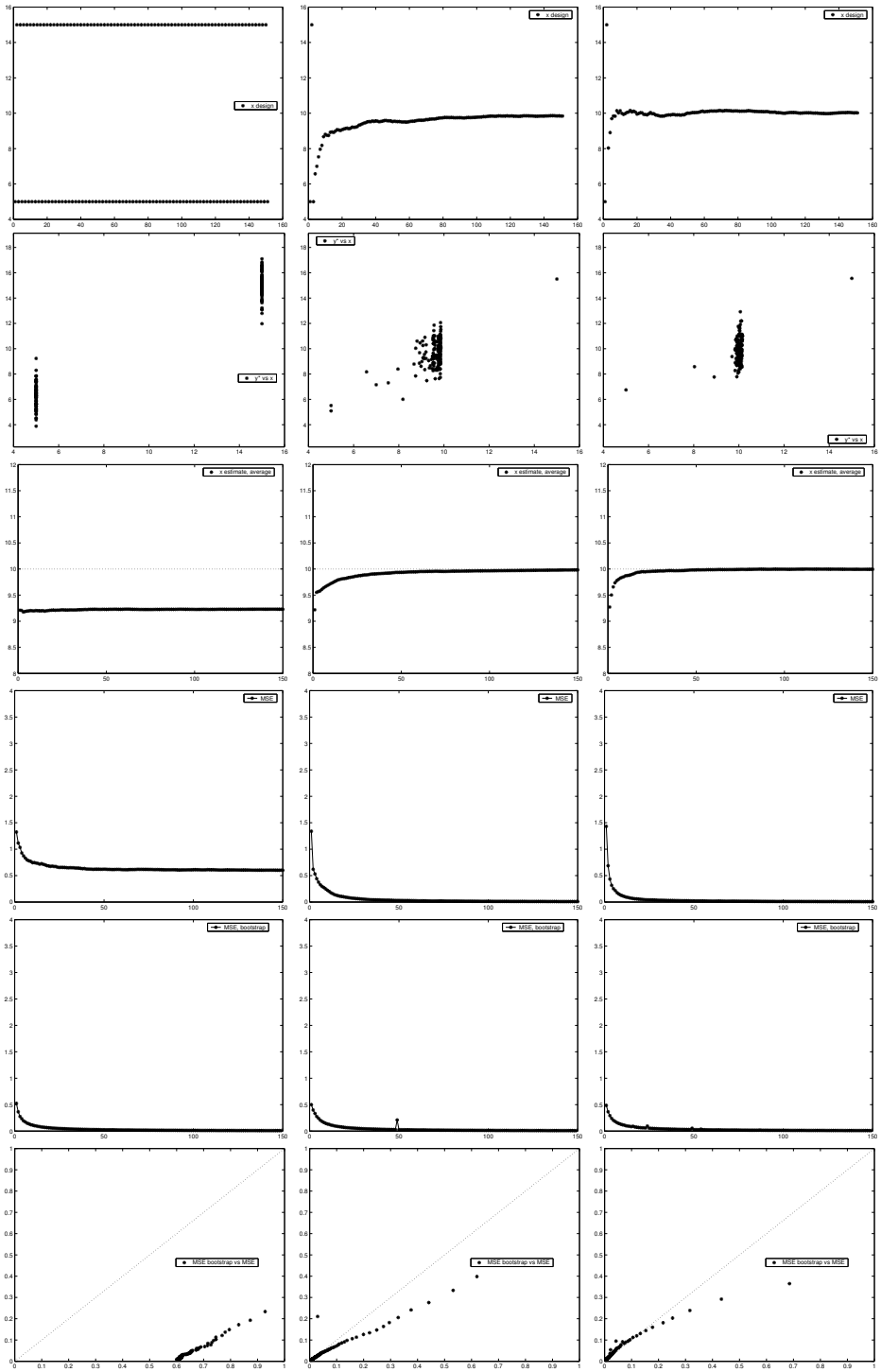


Figure 13: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the curved model/Design I, ie $(x_l, x_h) = (5, 15)$ and $\sigma_*^2 = 1.0$.

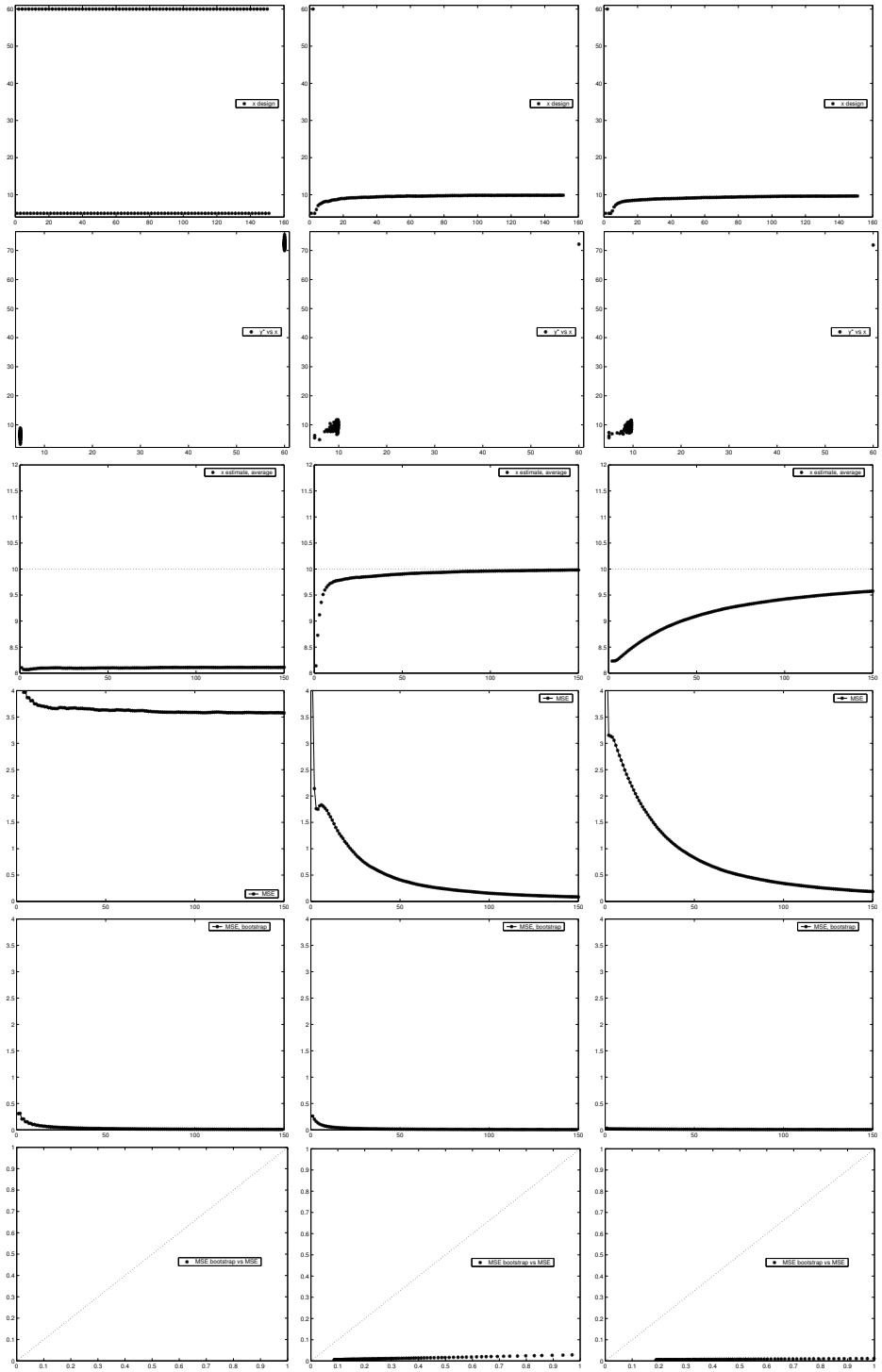


Figure 14: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the curved model/Design III, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 1.0$.

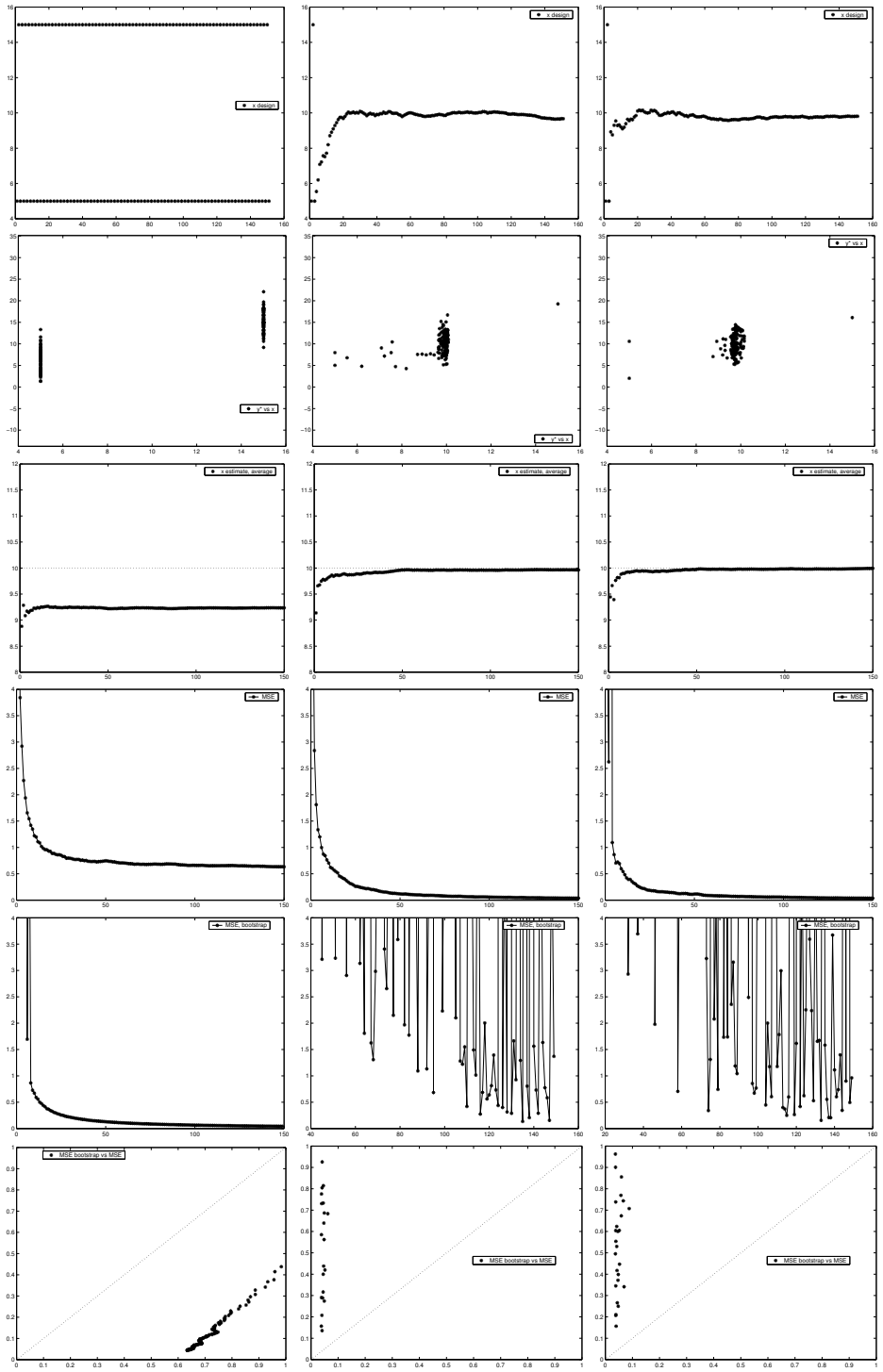


Figure 15: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the curved model/Design IV, ie $(x_l, x_h) = (5, 15)$ and $\sigma_*^2 = 5.0$.

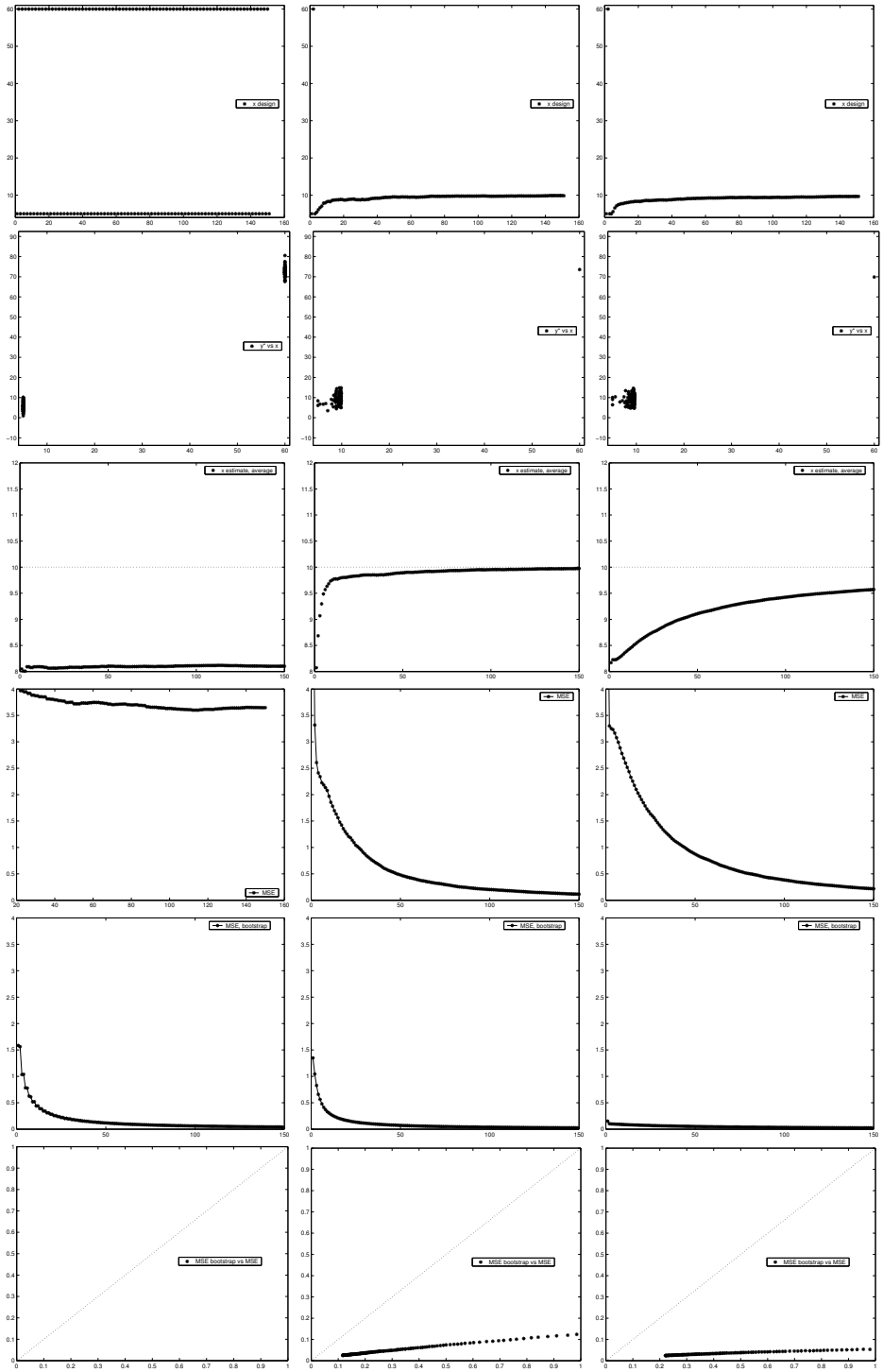


Figure 16: Simulation results of the D-, SA- and EI-algorithm with $\alpha = 1.0$ based on the curved model/Design IV, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 5.0$.

$\alpha = 0.51$ $\alpha = 0.75$ $\alpha = 1.0$

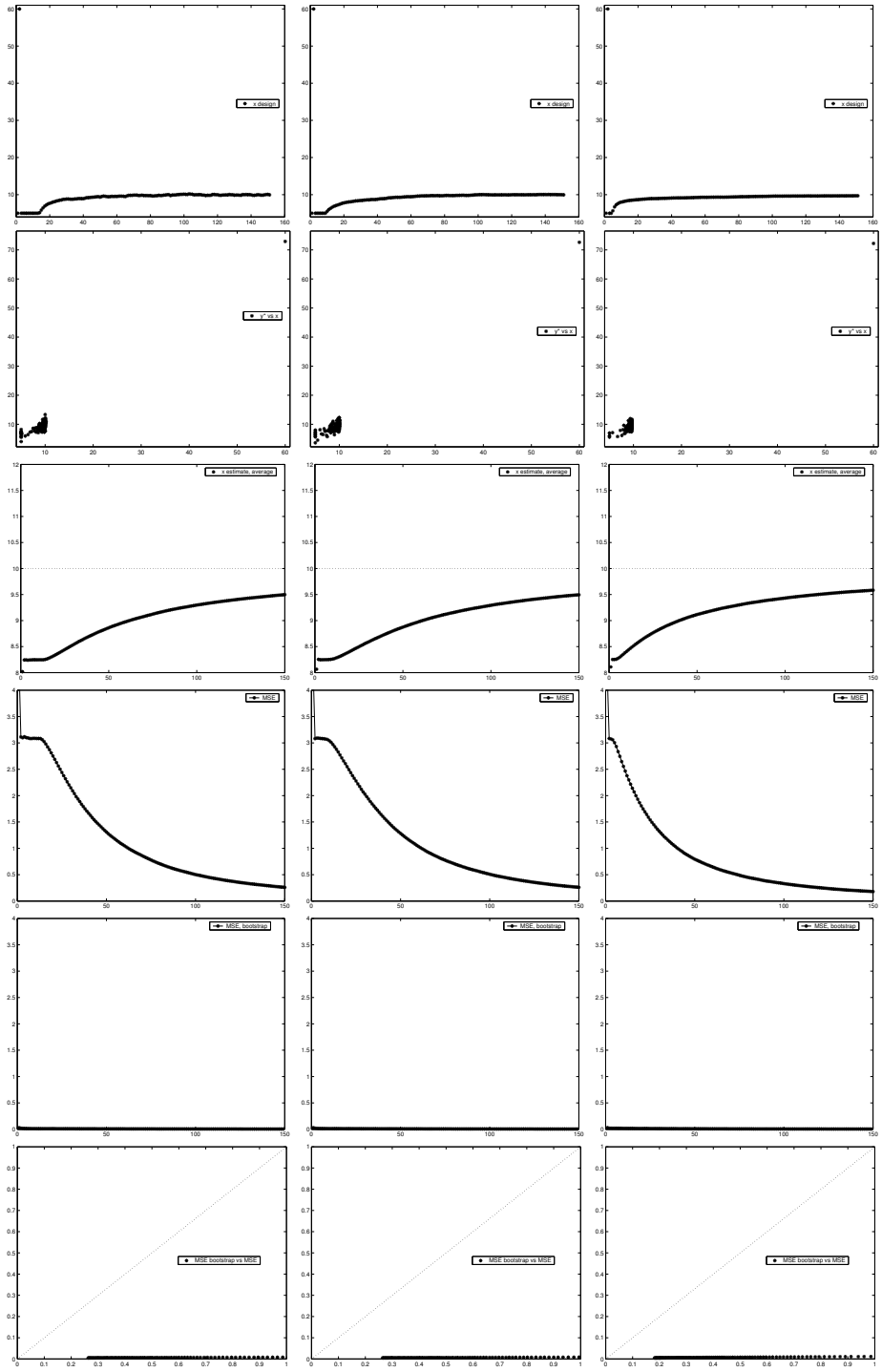


Figure 17: Simulation results of the EI-algorithm with varying α based on the curved model/Design III, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 1.0$.

$\alpha = 0.51$ $\alpha = 0.75$ $\alpha = 1.0$

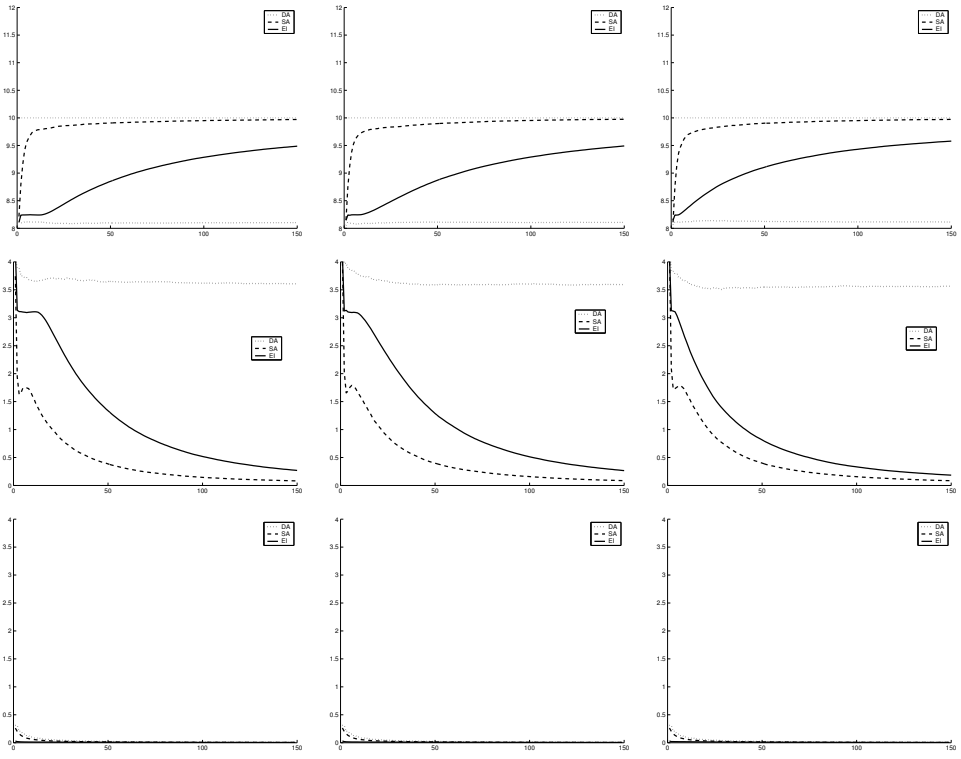


Figure 18: Comparison plots of D-, SA- and EI-algorithm with varying α based on the curved model/Design III, ie $(x_l, x_h) = (5, 60)$ and $\sigma_*^2 = 1.0$. The displays from top to bottom are estimates, estimated squared error and bootstrap estimate of estimation variance.

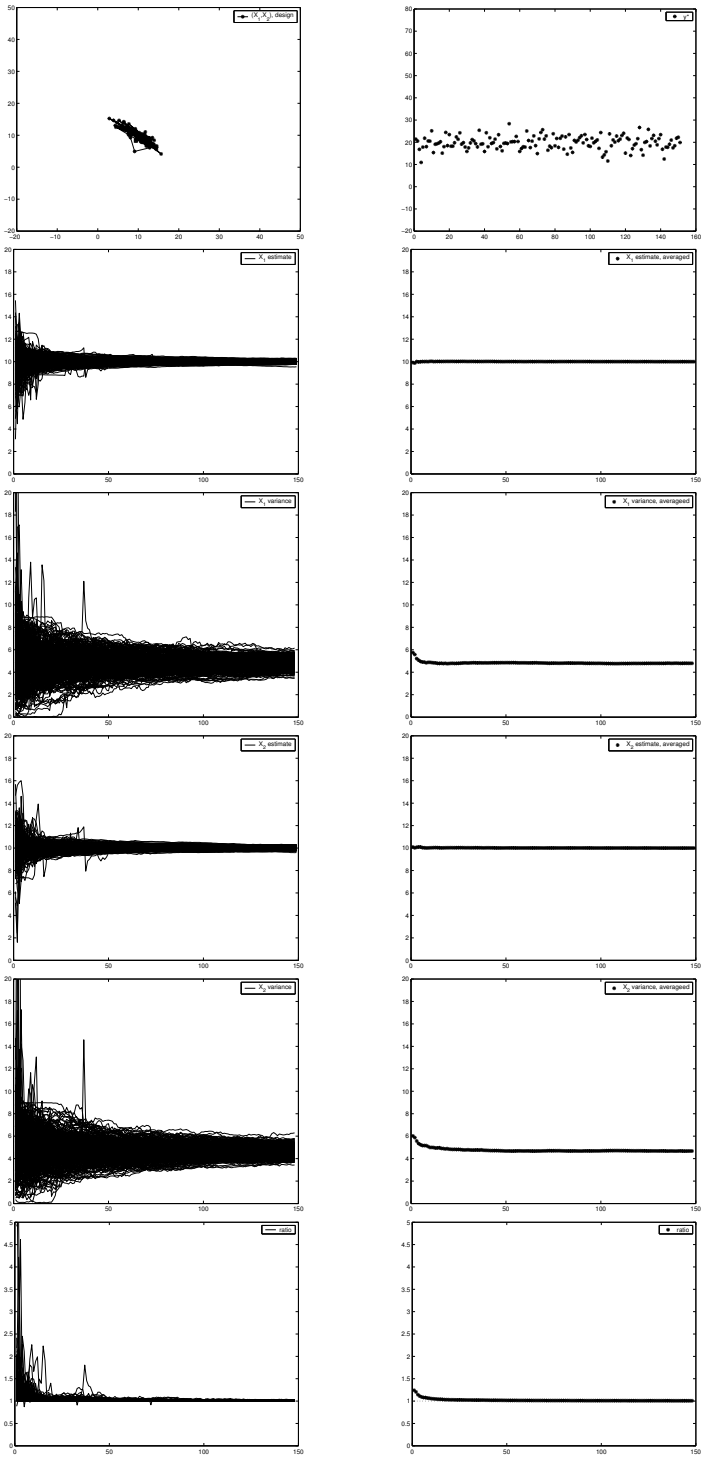


Figure 19: Simulation results of the experimental Bayesian inversion algorithm with $\alpha = 1.0$ based on the linear model/Design I, ie $\mu_X = (10, 10)$, $\sigma_X^2 = (9, 9)$.

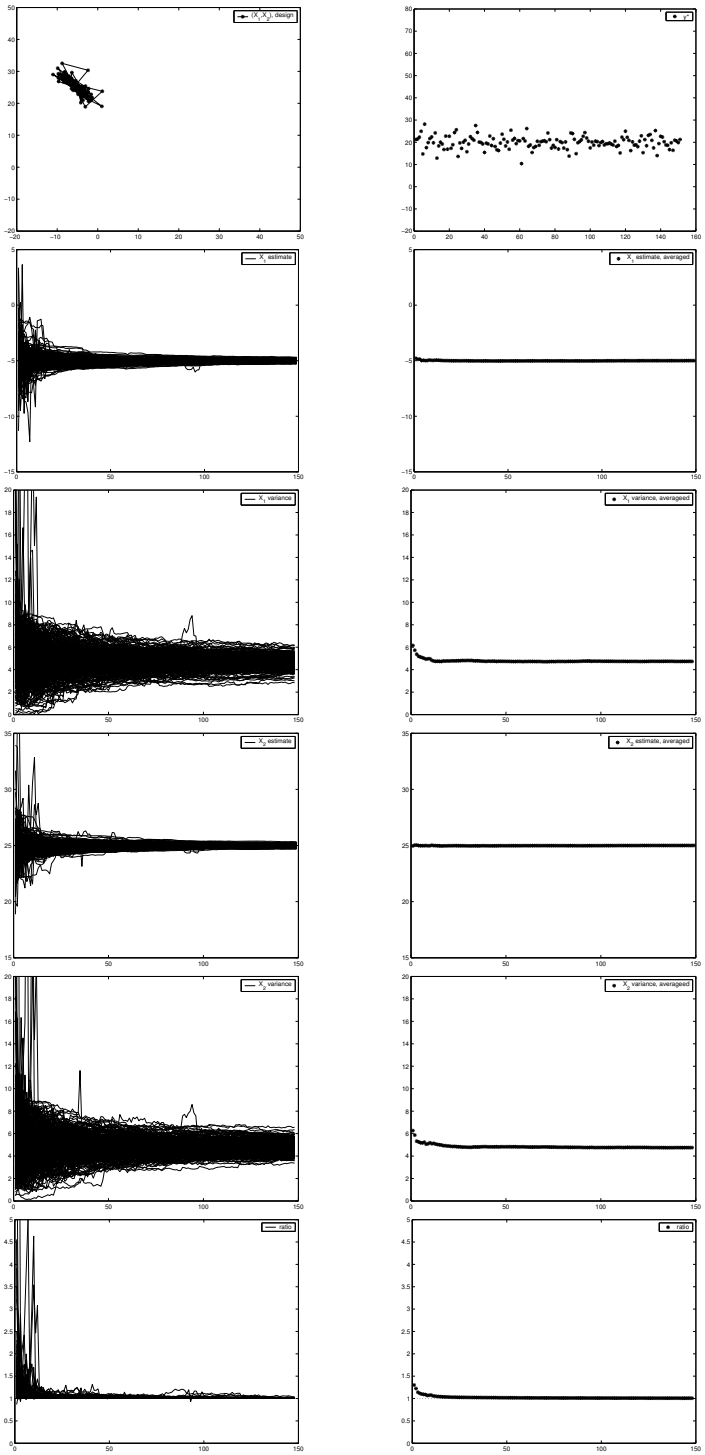


Figure 20: Simulation results of the experimental Bayesian inversion algorithm with $\alpha = 1.0$ based on the linear model /Design II, ie $\mu_X = (-5, 25)$, $\sigma_X^2 = (9, 9)$.

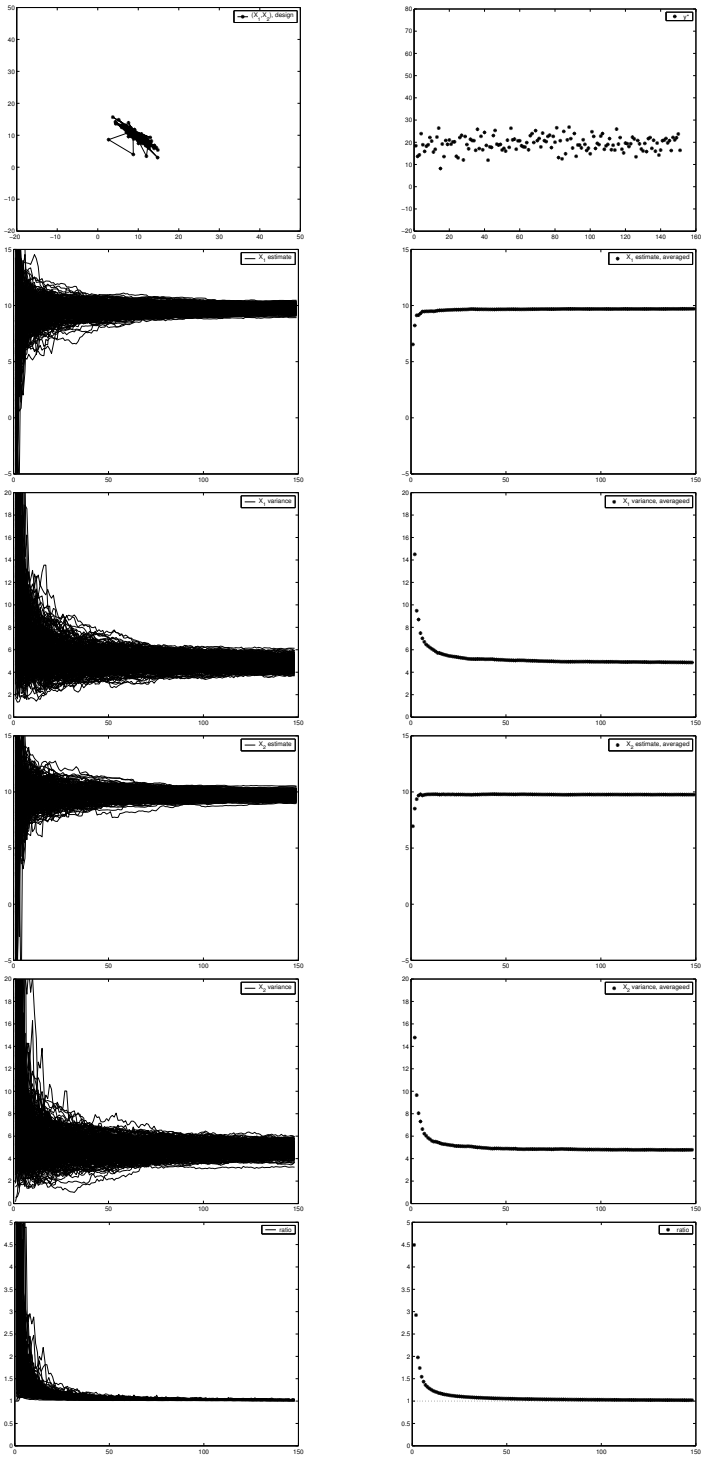


Figure 21: Simulation results of the experimental Bayesian inversion algorithm with $\alpha = 1.0$ based on the linear model /Design III, ie $\mu_X = (5, 5)$, $\sigma_X^2 = (9, 9)$.

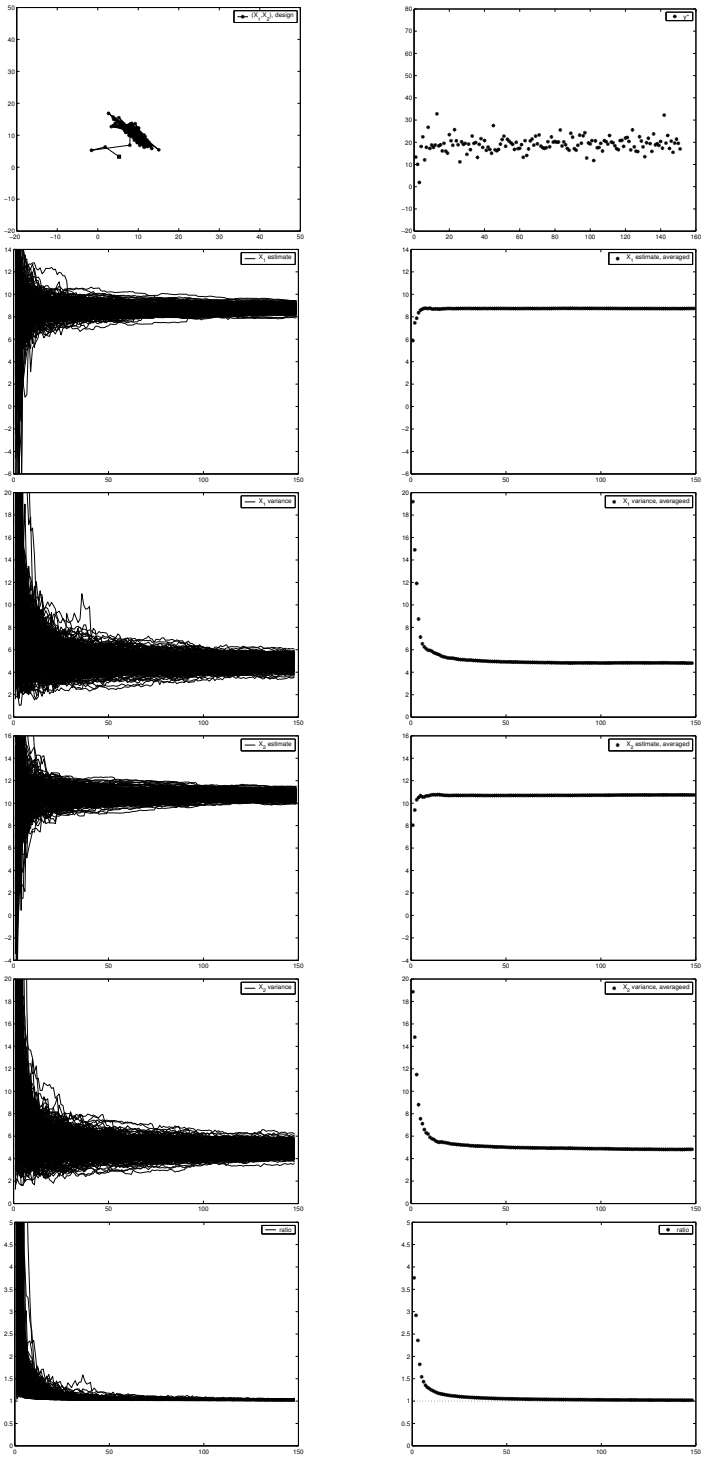


Figure 22: Simulation results of the experimental Bayesian inversion algorithm with $\alpha = 1.0$ based on the linear model /Design IV, ie $\mu_X = (4, 6)$, $\sigma_X^2 = (9, 9)$.

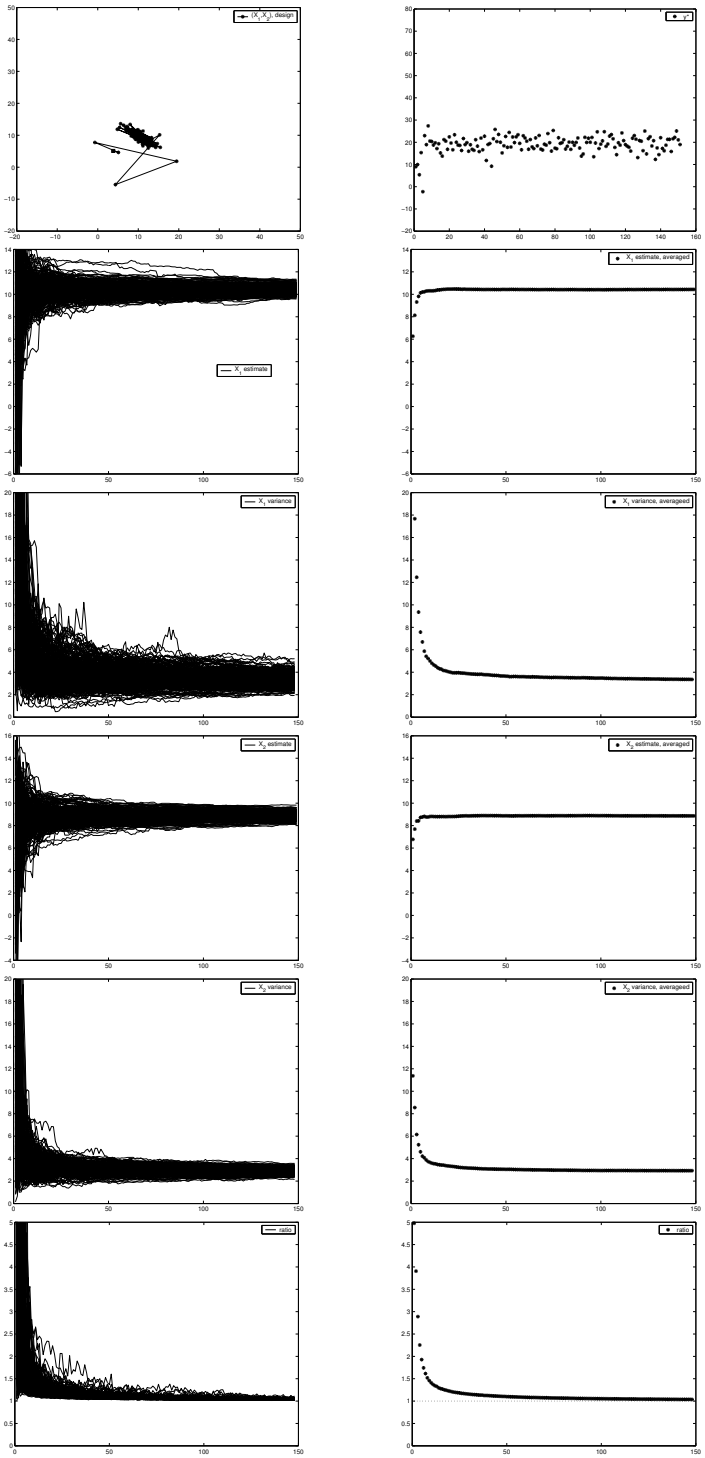


Figure 23: Simulation results of the experimental Bayesian inversion algorithm with $\alpha = 1.0$ based on the linear model /Design V, ie $\mu_X = (4, 6)$, $\sigma_X^2 = (9, 4)$.

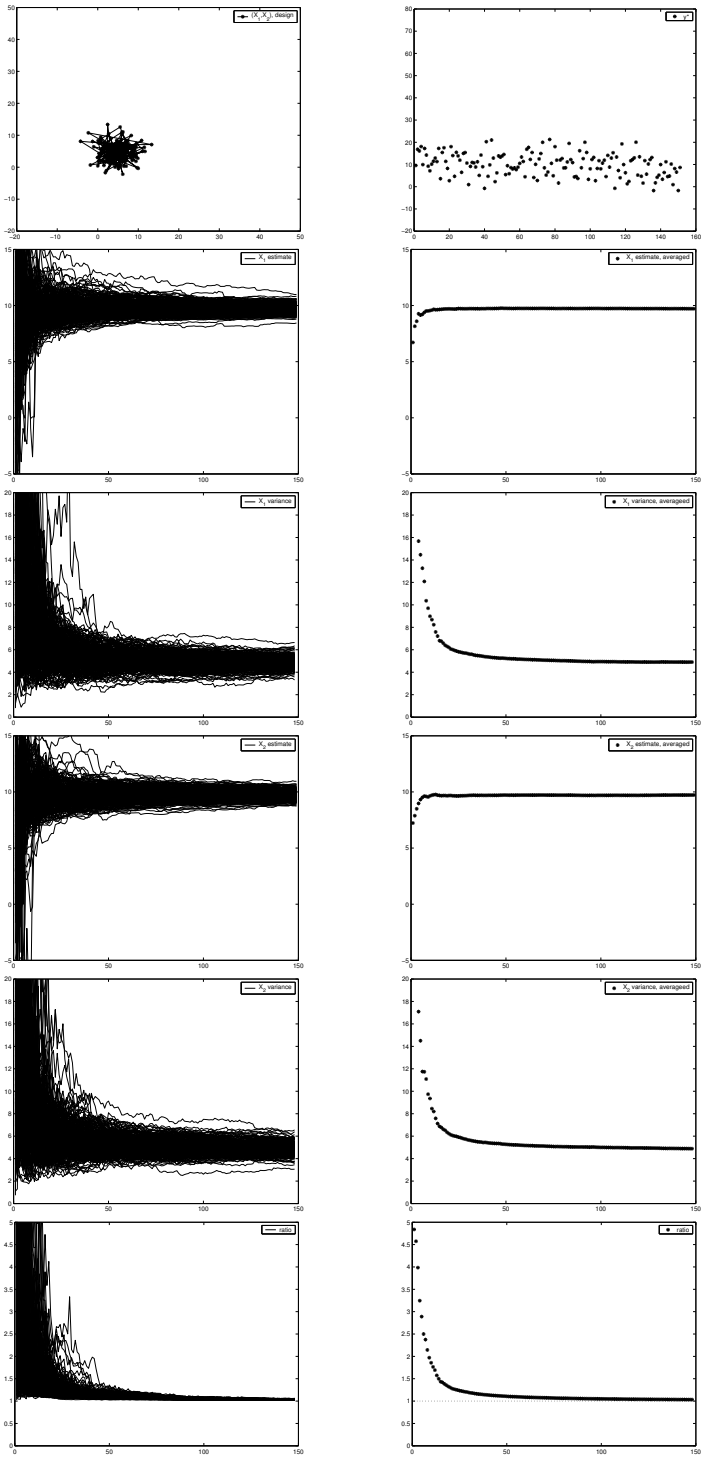


Figure 24: Simulation results of the prior-sampling experimental Bayesian inversion algorithm based on the linear model/Design III, ie $\mu_{_X} = (5, 5)$, $\sigma^2_{_X} = (9, 9)$.

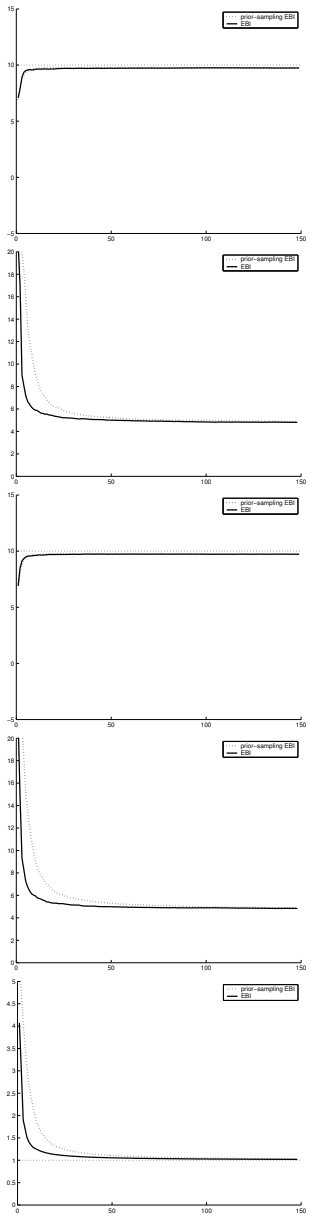


Figure 25: Comparison of the experimental Bayesian inversion algorithm with $\alpha = 1.0$ and the prior-sampling experimental Bayesian inversion algorithm based on the linear model/Design III, ie $\mu_X = (5, 5)$, $\sigma_X^2 = (9, 9)$. The two upper displays are average of estimate and bootstrap estimation variance in x_1-dimension while the display three and four are the corresponding ones in the x_2-dimension. Lower display is ratio of bootstrap estimation variance including uncertainty in $g(x)$ over estimation variance given $g(x)$.