Statistica Sinica **24** (2014), 197-210
doi:http://dx.doi.org/10.5705/ss.2011.291

THE STEPWISE RESPONSE REFINEMENT SCREENER (SRRS)

Frederick Kin Hing Phoa

Academia Sinica

Abstract: Supersaturated designs (SSDs) are useful in investigating a large number of factors with few experimental runs, particularly in screening experiments. The goal is to identify sparse but dominant active factors with small cost. In this paper, an analysis procedure called the Stepwise Response Refinement Screener (SRRS) method is proposed to screen important effects. Unlike the traditional approach that regresses factors with the response in every iteration, the response in each iteration of the SRRS is refined from the previous iteration using a selected potentially important factor. Analyses of two experiments using SSDs suggest that the SRRS method is able to retrieve similar results as do existing methods. Simulation studies show that, compared to existing methods in the literature, the SRRS method performs well in terms of the true model identification rate and the average model size.

Key words and phrases: Stepwise response refinement screener (SRRS), modified Akaike information criterion (mAIC), screening experiment, supersaturated design (SSD).

1. Introduction

As science and technology have advanced, investigators are becoming more interested in and capable of studying large-scale systems. Typically the initial stage of these systems contain a large number of potentially relevant factors, and it is common that the probing and studying of a large-scale system is expensive. To address these challenges, research in experimental design has lately focused on the class of supersaturated designs (SSDs) for their run-size economy. Under the condition of factor sparsity, or the principle of parsimony (Box and Meyer (1986)), the purpose of such experiments is to correctly identify the subset of those active factors that have significant impact on the response, so that the investigation can proceed via discarding inactive factors prior to follow-up experiments.

The construction of SSDs dates back to Satterthwaite (1959) and Booth and Cox (1962). The former suggested the use of random balanced designs and the latter proposed an algorithm to systematically construct SSDs. Many methods have been proposed for constructing them over the last 15 years. A comprehensive

list of early works can be found in Liu and Liu (2011) and Sun, Lin, and Liu (2011).

Traditionally, SSDs are employed primarily for screening main effects, discarding the possibility of interactions. Even the analysis considers main effects only, usual regression methods using all candidate factors cannot be used. More refined analysis methods were developed in recent years. Lin (1993) used stepwise regression for selecting active factors. Chipman, Hamada, and Wu (1997) proposed a Bayesian variable-selection approach for analyzing experiments with complex aliasing. Westfall, Young, and Lin (1998) proposed an error control skill in forward selection. Beattie, Fong, and Lin (2002) proposed a two-stage Bayesian model selection strategy for supersaturated experiments. Li and Lin (2002, 2003) proposed a method based on penalized least squares. Holcomb, Montgomery, and Carlyle (2003) proposed contrast-based methods. Lu and Wu (2004) proposed a modified stepwise selection based on an idea of staged dimensionality reduction. Zhang, Zhang, and Liu (2007) proposed a method based on partial least squares. Georgiou (2008) proposed an SVD principal regression method for SSDs.

Candes and Tao (2007) proposed the Dantzig selector (DS) and showed that it has remarkable properties under some conditions. The DS had been successfully used in biomedical imaging, analog to digital conversion, and sensor networks. Phoa, Pan, and Xu (2009) implemented the DS in practice. They introduced a graphical procedure using a profile plot in analyzing the results from the DS. In addition, they suggested an automatic variable selection procedure via a modified version of Akaike information criterion (AIC) to accompany the DS method. Traditionally, AIC is used for model selection. For linear models,

$$AIC = n \log\left(\frac{RSS}{n}\right) + 2p, \tag{1.1}$$

where $RSS = \sum_{i=1}^{n}(y_i - \hat{y}_i)^2$ is the residual sum of squares, n is the number of runs, and p is the number of parameters in the model. It is known that AIC tends to overfit the model when the sample size is small. Phoa, Pan, and Xu (2009) imposed a heavy penalty on the model complexity and proposed modified AIC for the automatic variable selection procedure of the DS method,

$$mAIC = n \log\left(\frac{RSS}{n}\right) + 2p^2. \tag{1.2}$$

The $mAIC$ typically chooses a smaller model than AIC.

The idea of a two-step procedure has been proposed before. For example, Fan and Lv (2008) proposed Sure Independence Screening (SIS) that essentially screens hte variables first and then performs model selection. Hwang and Hu

(2012) proposed Stepwise Paring down Variation (SPV) for identifying influential multi-factor interactions related to continuous response variables. In this paper, we introduce a variable selection approach via the Stepwise Response Refinement Screener (SRRS). The SRRS chooses the best subset of variables or active factors by two procedures: factor screening and model searching. Here "stepwise" only refers to the refinement of responses in every step - it is a forward selection procedure. Section 2 presents the procedural steps and Section 3 provides an explanation of them. To demonstrate the value of the SRRS method and compare it to other methods, the SSDs introduced by Lin (1993) and Rais et al. (2009) are considered in our simulation studies in Section 4. The last section gives some concluding remarks.

2. Analysis of SSDs via the Stepwise Response Refinement Screener (SRRS) Method

Consider a linear regression model $y = X\beta + \epsilon$, where y is an $n \times 1$ vector of observations, X is an $n \times k$ model matrix, β is a $k \times 1$ vector of unknown parameters, and ϵ is an $n \times 1$ vector of random errors. Assume that $\epsilon \sim N(\mathbf{0}, \sigma^2 I_n)$. In addition, X is assumed to be supersaturated, $n < k$. We let m be the number of potentially important effects (PIEs) and S_{inf} be the influential set of PIEs found in the process.

The Stepwise Response Refinement Screener (SRRS) proceeds as follows.

I. SRRS–Factor Screening

Step 1. Standardize data so that y_0 has mean 0 and the columns of X have equal lengths.

Step 2. Compute the correlation $\rho(X_i, y_0)$ for all factors X_i, $i = 1, \ldots, k$.

Step 3. Choose E_0 such that $|\rho(E_0, y_0)| = \max_{X_i} |\rho(X_i, y_0)|$ and include E_0 as the first PIE in S_{inf}.

Step 4. Obtain the estimate β_{E_0} by regressing y_0 on E_0. Unless specified by expert opinion, take the threshold of noise level γ to be approximately $5\% - 10\%$ of β_{E_0}.

Step 5. For the next m PIEs E_j, $j = 1, \ldots, m$, $m < n - 2$,

(a) compute the refined response $y_j = y_{j-1} - E_{j-1}\beta_{E_{j-1}}$;

(b) compute the marginal correlation $\rho(X_i, y_j)$ for all X_i, $i = 1, \ldots, k$;

(c) choose T_j such that $|\rho(T_j, y_j)| = \max_{X_i} |\rho(X_i, y_j)|$;

(d) obtain the estimate β_{T_j} by regressing y_j on $E_0, \ldots, E_{j-1}, T_j$;

(e) if $|\beta_{T_j}| \geq \gamma$ and T_j has not been included in S_{inf}, put $E_j = T_j$ and include it in S_{inf};

(f) repeat (a) to (e) up to m^{th} step, where $E_j = E_m$ is not included in S_{inf}, m determined by either $m < n-2$ or the threshold condition $|\beta_{T_j}| \geq \gamma$, or both.

II. SRRS–Model Searching

Step 6. Perform an all-subset search for all E_j, from models with one to m factors, where m is minimum of $n/3$ and the number of E_j in S_{inf}.

Step 7. Compute $mAIC$ for each model and choose the final model as the one with smallest $mAIC$; all E_j included in the final model are considered to be significant to the response y_0.

3. Main Idea of the SRRS Method

Step 1 is a standard normalization on the response y_0 and the factors X. The factor with the highest marginal correlation to y_0 is identified as the first PIE and denoted as E_0. Once E_0 is included in S_{inf}, a model between y_0 and E_0 is built and the slope estimate β_{E_0} is obtained; it is used for defining the threshold of noise level γ.

Here γ is a threshold between signal and noise and a relatively small γ should be chosen. One can choose γ according to information on the magnitude of effects or noise. For example, Phoa, Pan, and Xu (2009) suggested that γ is approximately 10% of the maxmimum absolute estimates in their simulation study. It is recommended that the procedure is repeated with a few choices of γ. When the signal to noise ratio is large, the choice of γ is not crucial. Generally speaking, we choose γ to be approximately $5\% - 10\%$ of $|\beta_{E_0}|$ in our examples and simulation studies. Although $|\beta_{E_0}|$ may not be the maximum slope estimate in some cases, it is conservative to set a slightly smaller γ so that one or two more factors are considered as PIEs.

Before starting to search for the next PIE, the response has to be refined so that the correlation between the newly refined response and E_0 is close to 0. This is what Step 5(a) does. The first refined response, y_1, is the difference between the original response y_0 and the portion of magnitude of the first PIE $E_0\beta_{E_0}$. The search repeats the same procedure for the next m PIEs, and $E_1, \ldots, E_{m-1}$ are included in S_{inf}. The marginal correlation between y_j and T_{j-1} is zero, and the marginal correlations between y_j and all other factors, including those that have been included in S_{inf}, are compared in Step 5(b). The magnitudes of these marginal correlations consist of: (i) some middle to high values that indicate these factors still have possibilities to be PIEs after j refinements, and (ii) some close-to-zero values that indicate these factors do not have impact on the response.

There are two criteria that can stop the search. The first is the number of PIEs in S_{inf}; It has to be at most the number of runs minus two so that one can estimate all PIEs, the intercept, and the residual of the model. The second criterion is related to the magnitude of the slope estimate, magnitudes lower than

Table 1. A Two-level SSD (Lin (1993)).

Run	Factors																							Response
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	Y
1	+	+	+	−	−	−	+	+	+	+	+	−	−	−	+	+	−	−	+	−	−	−	+	133
2	+	−	−	−	−	−	+	+	+	−	−	−	+	+	+	−	+	−	−	+	+	−	−	62
3	+	+	−	+	+	−	−	−	−	+	−	+	+	+	+	+	−	−	−	−	+	+	−	45
4	+	+	−	+	−	+	−	−	−	+	+	−	−	+	+	−	+	+	+	−	−	−	−	52
5	−	−	+	+	+	+	−	+	+	−	−	−	−	+	+	+	−	−	+	−	+	+	+	56
6	−	−	+	+	+	+	+	−	+	+	+	−	+	+	−	+	+	+	+	+	+	−	−	47
7	−	−	−	−	+	−	−	+	−	+	−	+	+	−	+	+	+	+	+	+	−	−	+	88
8	−	+	+	−	−	+	−	+	−	+	−	−	−	−	−	−	−	+	−	+	+	+	−	193
9	−	−	−	−	−	+	+	−	−	−	+	+	−	+	−	+	+	−	−	−	−	+	+	32
10	+	+	+	+	−	+	+	+	−	−	−	+	+	+	−	+	−	+	−	+	−	−	+	53
11	−	+	−	+	+	−	−	+	+	−	+	−	+	−	−	−	+	+	−	−	−	+	+	276
12	+	−	−	−	+	+	+	−	+	+	+	+	−	−	+	−	−	+	−	+	+	+	+	145
13	+	+	+	+	+	−	+	−	+	−	−	+	−	−	−	−	+	−	+	+	−	+	−	130
14	−	−	+	−	−	−	−	−	−	−	+	+	+	−	−	−	−	−	+	−	+	−	−	127

γ are considered as noise. If E_j is chosen in Step 5(c) but $|\beta_{E_j}|$ is found to be smaller than γ, the search stops and all remaining factors are considered as noise.

Step 6 constructs reduced models with all possible combinations of factors - m one-factor models, C_2^m two-factor models, and so on. Instead of building all possible reduced models, Marley and Woods (2010) suggested that the number of runs is at least three times the number of active factors. We follow their guideline so the maximum number of active factors in the reduced models is less than or equal to $n/3$, unless the number of PIEs is smaller.

Instead of using the traditional AIC, $mAIC$ (Phoa, Pan, and Xu (2009)) is the criterion for the best among all reduced models. Phoa, Pan, and Xu (2009) showed that $mAIC$ performed better than both the traditional AIC and another modified version $cAIC$ (Hurvich and Tsai (1989)) when the model is supersaturated. We suggest that all PIEs included in the reduced model with the smallest $mAIC$ be considered to have significant impact even if some PIEs are not significant based on their p-values.

4. Two Illustrative Examples

We illustrate the analysis of SSDs via the SRRS method.

Example 1. Consider the SSD of Lin (1993). The original dataset has 24 factors but factors 13 and 16 are identical. As Beattie, Fong, and Lin (2002) and Phoa, Pan, and Xu (2009) did, we delete factor 13 and rename factors 14–24 as 13–23. The design matrix and response data are given in Table 1.

Table 2. Factor Screening of Two-Level SSD of Lin (1993) via SRRS.

m	PIE	Marginal Correlation	$\lvert\beta\rvert$	Continue or Stop
0	X_{14}	-0.7948	53.21	Continue
1	X_{12}	-0.5370	22.27	Continue
2	X_{19}	-0.6751	24.78	Continue
3	X_4	-0.5730	22.12	Continue
	X_{14}	-0.7326	17.26	Continue
4	X_{10}	-0.5904	9.40	Continue
5	X_{11}	0.6569	8.16	Continue
6	X_7	-0.6989	7.78	Continue
	X_{19}	-0.7337	4.68	Stop
PIEs in S_{inf} after Factor Screening: X_4, X_7, X_{10}, X_{11}, X_{12}, X_{14}, X_{19}.				

In the factor screening procedure, the first PIE identified is X_{14}. A regression model between y_0 and X_{14} is built and one finds $|\beta_{X_{14}}| = 53.21$. We set the threshold $\gamma = 5$, about 10% of $|\beta_{X_{14}}|$.

The second PIE is X_{12}. A regression model between y_1 and X_{12}, X_{14} is built and here $|\beta_{X_{12}}| = 22.27 > \gamma$; X_{12} is included in the influential set S_{inf} together with X_{14}.

Table 2 shows all steps of the process of factor screening. The first two columns are the number and the name of PIEs identified, the third column shows the marginal correlations between the PIEs and the refined responses, and the fourth column shows the absolute magnitude of slope estimates in the models between the PIEs and the refined responses.

After the fourth PIE X_4 is identified, the factor that has the highest absolute marginal correlation to the refined response is X_{14} again. This situation happens due to the collinearity among factors. A similar situation happens after the seventh PIE X_7 is identified, where X_{19} has the highest absolute marginal correlation to the refined response. Here, since $|\beta_{X_{19}}| = 4.68 < \gamma$, the search stops and the influential set S_{inf} that consists of seven PIEs is fixed.

Since there are 14 observations, the maximum number of active factors is 5 if the guideline of Marley and Woods (2010) is followed. Then there are 119 reduced models built in Step 6, including 7 one-factor models, 21 two-factor models, 35 three-factor models, 35 four-factor models and 21 five-factor models. Comparing $mAIC$, the one-factor model with X_{14} has the lowest $mAIC$. Thus the SRRS method suggests that only X_{14} has significant impact to the response y_0.

The same data were previously analyzed by several authors. Beattie, Fong, and Lin (2002) compared several model selection methods and identified X_{14} as the only important effect in every method. By using the Dantzig selector,

Phoa, Pan, and Xu (2009) also suggested that all other effects except X_{14} are noisy and the magnitudes are small in the profile plot. Both Li and Lin (2003) and Zhang, Zhang, and Liu (2007) suggested X_{14}, X_{12}, X_{19}, and X_4 as active factors. Westfall, Young, and Lin (1998) added an extra active factor X_{11} to the previous list, among which X_{14} is the only significant variable at 5% significance level. A regression model between the original response and X_{14} shows that the p-value of X_{14} is 0.000681, X_{14} is highly significant to the response.

The analysis via SRRS matches the conclusion in Beattie, Fong, and Lin (2002). Although all other PIEs except X_{14} in either our work or any previous works are barely significant to the response, we agree with Abraham, Chipman, and Vijayan (1999) that it is still not clear what the active factors are.

Example 2. Consider the SSD of Rais et al. (2009). The dataset has 31 factors and 18 runs. The design matrix and response data are given in Table 3.

In the factor screening procedure, the first PIE identified is U_{28}. A regression model between y_0 and U_{28} is built and one finds $|\beta_{U_{28}}| = 8.66$. We set the threshold $\gamma = 0.85$, about 10% of $\beta_{U_{28}}$.

Table 4 shows every step of the factor screening that results in thirteen PIEs. With 18 observations, the maximum number of active factors is suggested to be 6, and 4095 reduced models are built and their $mAIC$ compared. The two-factors model with U_{24} and U_{27} has the lowest $mAIC = 90.36$. A regression model between the original response and U_{24} and U_{27} shows that the p-values of U_{24} and U_{27} are 0.00305 and 0.00731, respectively, U_{24} and U_{27} are highly significant.

Rais et al. (2009) suggested nine significant factors, U_{13}, U_{18}, U_{19}, U_{20}, U_{24}, U_{27}, U_{28}, U_{29} and U_{30}. Their model had $R^2 = 0.9879$, compared to the SRRS $R^2 = 0.5124$. Traditional wisdom suggests that AIC is a better measure on the goodness of fit because it penalizes the number of parameters, but Phoa, Pan, and Xu (2009) showed that AIC may still overfit under the condition of a SSD. The $mAIC$ of Rais et al. (2009) model is 180.09, which is almost double of the $mAIC$ of the model suggested by the SRRS.

It is difficult to say whether Rais et al. (2009) model or the model suggested by SRRS is better in general, but both analyses have U_{24} and U_{27} significant. Some follow-up experiments or analysis should be done for further investigation.

5. Simulation Studies

To judge the value of the SRRS method, we compared its performance with that of five different approaches: SSVS, the Bayesian variable selection procedure proposed by George and McMulloch (1993) and extended for SSDs by Chipman, Hamada, and Wu (1997); SSVS/IBF, the two stage Bayesian procedure of Beattie, Fong, and Lin (2002); SCAD, the penalized least squares approach proposed

Table 3. A Two-level SSD (Rais et al. (2009)).

Run	Factors															
	U_1	U_2	U_3	U_4	U_5	U_6	U_7	U_8	U_9	U_{10}	U_{11}	U_{12}	U_{13}	U_{14}	U_{15}	U_{16}
1	−	+	−	+	+	+	−	−	−	+	+	+	+	+	−	+
2	+	−	−	+	−	+	+	+	−	−	−	+	+	+	+	+
3	−	+	−	−	+	−	+	+	+	−	−	−	+	+	+	+
4	+	+	−	−	+	−	−	+	−	+	+	+	−	−	−	+
5	+	−	+	+	−	−	+	−	−	+	−	+	+	+	−	−
6	+	−	+	−	+	+	−	−	+	−	−	+	−	+	+	+
7	−	+	−	+	−	+	+	−	−	+	−	−	+	−	+	+
8	−	−	+	−	+	−	+	+	−	−	+	−	−	+	−	+
9	−	−	−	+	−	+	−	+	+	−	−	+	−	−	+	−
10	+	−	−	−	−	+	−	+	−	+	+	−	−	+	−	−
11	−	+	−	−	−	−	+	−	+	−	+	+	−	−	+	−
12	+	+	+	−	−	+	−	−	−	−	+	−	+	−	+	+
13	+	+	+	+	+	−	+	+	+	−	−	+	−	−	−	−
14	−	+	+	+	+	+	−	+	+	+	−	−	+	−	−	−
15	−	−	+	+	+	+	+	−	+	+	+	−	−	+	−	−
16	+	+	+	−	−	−	+	+	+	+	+	−	+	+	+	−
17	+	−	+	+	+	−	−	−	+	+	+	+	+	−	+	+
18	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−

Run	Factors															
	U_{17}	U_{18}	U_{19}	U_{20}	U_{21}	U_{22}	U_{23}	U_{24}	U_{25}	U_{26}	U_{27}	U_{28}	U_{29}	U_{30}	U_{31}	Y
1	+	+	−	−	+	−	−	−	−	+	−	+	−	+	+	76.2
2	−	+	+	+	−	−	+	−	−	−	−	+	−	+	−	82.6
3	+	−	+	+	+	−	−	+	−	−	−	−	+	−	+	99.4
4	+	+	+	+	−	+	+	+	−	−	+	−	−	−	−	80.5
5	−	+	+	+	+	+	−	+	+	+	−	−	+	−	−	103.5
6	−	−	−	+	+	+	+	+	−	+	+	+	−	−	+	52.1
7	+	−	−	−	+	+	+	+	+	−	+	+	+	−	−	73.8
8	+	+	−	−	−	+	+	+	+	+	−	+	+	+	−	89.8
9	+	+	+	−	−	−	+	+	+	+	+	−	+	+	+	100.7
10	+	−	+	+	+	−	−	−	+	+	+	+	+	−	+	59.8
11	−	+	−	+	+	+	−	−	−	+	+	+	+	+	−	62.8
12	−	−	+	−	−	+	−	+	+	+	−	−	−	+	+	95.0
13	+	−	+	−	+	+	−	−	+	−	−	+	−	+	+	74.9
14	−	+	−	+	−	+	+	−	−	+	−	−	+	−	+	84.4
15	−	−	+	−	+	−	+	+	−	−	+	−	−	+	−	86.7
16	−	+	−	−	−	−	+	−	+	−	+	+	−	−	+	58.3
17	+	−	−	+	−	−	−	−	+	−	+	−	+	+	−	71.1
18	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	64.8

by Li and Lin (2003); PLSVS, the partial least square regression technique of Zhang, Zhang, and Liu (2007); SVDPR, the singular value decomposition principal regression of Georgiou (2008). We also consider DS, the l_1-regularization

Table 4. Factor Screening of Two-Level SSD of Rais et al. (2009) via SRRS.

m	PIE	Marginal Correlation	$\lvert\beta\rvert$	Continue or Stop
0	U_{28}	−0.5763	8.66	Continue
1	U_{27}	−0.4864	6.05	Continue
2	U_{24}	0.5530	7.10	Continue
3	U_{30}	0.5093	4.25	Continue
4	U_8	0.5176	3.89	Continue
5	U_4	0.4888	3.99	Continue
6	U_5	−0.5141	2.94	Continue
7	U_7	0.4732	1.65	Continue
8	U_{29}	0.4762	2.44	Continue
9	U_{14}	0.5016	1.45	Continue
10	U_2	0.6197	3.23	Continue
	U_{14}	0.5486	2.03	Continue
11	U_{12}	0.4884	1.01	Continue
12	U_{11}	0.5273	1.81	Continue
	U_3	−0.5660	0.85	Stop
PIEs in S_{inf} after Factor Screening: U_4, U_7, U_{10}, U_{11}, U_{12}, U_{14}, U_{19}.				

of Phoa, Pan, and Xu (2009).

Example 3. We consider the simulation used in Li and Lin (2003), Zhang, Zhang, and Liu (2007), and Phoa, Pan, and Xu (2009). The design is a 14-run SSD with 23 factors used by Lin (1993) to analyze the popular dataset first reported in Williams (1968). Three regression models are considered:

Model I: $y = 10x_1 + \epsilon$;

Model II: $y = -15x_1 + 8x_5 - 2x_9 + \epsilon$; and

Model III: $y = -15x_1 + 12x_5 - 8x_9 + 6x_{13} - 2x_{16} + \epsilon$.

We generated data from the linear model where X is the SSDs of Lin (1993) given in Table 1 and $\epsilon \sim N(0, 1)$ is the random error. In our method, we chose thresholds of noise level $\gamma = 1$ and $\gamma = 0.75$. The former is approximately 10% (case I) or 6.7% (cases II and III) of $\max|\beta_i|$ while the latter is approximately equal to 7.5% (case I) or 5% (cases II and III) of $\max|\beta_i|$. Each simulations was run 1000 times. Table 5 compares performances; in it, "TMIR" stands for True Model Identified Rate, "SEIR" stands for Smallest Effect Identified Rate, and "Median" and "Mean" are the median and mean sizes of the models.

The SRRS method identifies the true model with the highest probabilities among the six methods. In case I, the SRRS method shares 100% perfect identification rates with the SCAD, PLSVS, and DS methods in identifying the smallest effect. In case II, the performance of the SRRS method is as good as that of the

Table 5. Comparison of simulation results in Example 3.

Case	Method	TMIR	SEIR	Median	Mean
I	SSVS(1/10,500)	40.5%	99.0%	2	3.1
	SSVS(1/10,500)/IBF	61.0%	98.0%	1	2.5
	SCAD	75.6%	100.0%	1	1.7
	PLSVS (m=1)	61.0%	100.0%	1	1.5
	SVDPR (CV)	94.2%	100.0%	1	1.07
	SVDPR ($a_{opt} = 0$)	98.3%	100.0%	1	1.02
	DS ($\gamma = 1$)	99.4%	100.0%	1	1.0
	SRRS ($\gamma = 1$)	99.8%	100.0%	1	1.0
	SRRS ($\gamma = 0.75$)	90.7%	100.0%	1	1.0
II	SSVS(1/10,500)	8.6%	30.0%	3	4.7
	SSVS(1/10,500)/IBF	8.0%	28.0%	3	4.2
	SCAD	75.6%	98.5%	3	3.3
	PLSVS (m=1)	76.4%	100.0%	3	3.3
	DS ($\gamma = 1$)	84.4%	85.3%	3	2.9
	SRRS ($\gamma = 1$)	84.2%	85.2%	3	2.9
	SRRS ($\gamma = 0.75$)	89.8%	92.5%	3	3.0
III	SSVS(1/10,500)	36.4%	84.0%	6	8.0
	SSVS(1/10,500)/IBF	40.7%	75.0%	5	5.6
	SCAD	69.7%	99.4%	5	5.4
	PLSVS (m=1)	73.6%	95.0%	5	5.2
	DS ($\gamma = 1$)	79.1%	91.2%	5	5.1
	SRRS ($\gamma = 1$)	95.3%	95.3%	5	5.0
	SRRS ($\gamma = 0.75$)	96.6%	96.6%	5	5.0

DS method in terms of the ability to identify the true model, the ability to identify the smallest effects, and the average model sizes for $\gamma = 1$, and the performance further improves for $\gamma = 0.75$. In case III, the performance between the DS method and the SRRS is clearly distinguishable, and the SRRS method performs the best among all six methods. In this sense the SRRS method seems to be more efficient, no matter which γ is chosen. Notice that the results of SVDPR are not good in cases II and III and thus we do not include them in the table.

The SRRS method is excellent in identifying active factors in Cases I and III. The performance is less effective in Case II mainly because the model consisting of factors 1, 5 and 12 explains the response as well as the true model. The randomly assigned error ϵ has a great effect. In fact, the missing 15% (when $\gamma = 1$) and 10% (when $\gamma = 0.75$) in SEIR is mainly due to the mis-identification of the significance of factor 12 instead of factor 9 in both the DS method and the SRRS method.

This example lends insight on the choice of γ. When the signal-to-noise ratio is large (case I), a relatively large γ helps in eliminating the potential falsely-

Table 6. Summary of simulation results in Example 4.

Case		Min	1st Quartile	Median	Mean	3rd Quartile	Max
I	TMIR	97%	99%	100%	99.49%	100%	100%
	Size	1.00	1.00	1.00	1.005	1.01	1.03
II	TMIR	96%	99%	100%	99.38%	100%	100%
	Size	2.00	2.00	2.00	2.006	2.01	2.04
III	TMIR	5%	99%	100%	96.96%	100%	100%
	Size	2.05	3.00	3.00	2.981	3.01	3.31
IV	TMIR	0%	93%	99%	84.33%	100%	100%
	Size	1.00	3.92	4.00	3.741	4.00	4.84
V	TMIR	0%	20%	86%	64.22%	100%	100%
	Size	1.00	3.39	4.77	4.145	5.00	5.00

identified PIEs in the factor screening procedure. On the other hands, when the signal-to-noise ratio is small (cases II and III), a relatively small γ includes a relatively larger number of PIEs in order to prevent the false elimination of the true active factors with small magnitude. The proper balance likely comes from experts or experience.

Example 4. We generated data from the SSD in Example 1. Since there are only 14 observations, the maximum number of active factors is 5, and we consider five cases for *beta*. There are i active factors for case i, $1 \leq i \leq 5$. For each case, we generated 500 models where the selection of active factors was random, the signs of the active factors were randomly selected to be positive or negative, and the magnitudes were randomly selected from 2 to 10. For each model, we generated data 100 times and obtain the True Model Identified Rate (TMIR) and the average model size. In the simulations we took $\gamma = 1$, it is approximately equal to 10% of $\max |\beta_i|$. Table 6 gives the summary statistics.

The SRRS method is very effective in identifying one and two active factors; the TMIR in both cases are at least 96% and only a few cases that have average model sizes slightly higher than the true numbers of active factors. The performance of the method slightly decreases in identifying three or four active factors. In case III, more than 95% of the 500 TMIR are $\geq 80\%$, and in one specific case, the TMIR is only 5%. In case IV, about 80% of the 500 TMIR are $\geq 80\%$. The situation becomes worse for case V, although more than half of the 500 TMIR are $\geq 80\%$.

We have rerun other simulations found in the literature. For example, the simulations for ten models in Georgiou (2008). The performance of the SRRS method was excellent, the true model identified rates above 99% in models 2, 3, 5, 6, 7 and 10 there.

6. Concluding Remarks

SSDs are useful for investigating in a large number of factors with few experimental runs, particularly in screening experiments. We introduce the Stepwise Response Refinement Screener (SRRS) method for use with SSDs. Simulations and the analysis of experiments suggested that the SRRS method performs better than existing methods, though not in all situations. The R function of the SRRS is available by email request from the author, and the standalone program for the SRRS will be available soon.

Once a suggested set of significant factors is found, a follow-up experiment is needed for validating the results. Then it is more economical and efficient to use nonregular fractional factorial designs than full factorial designs. A detailed review of nonregular fractional factorial designs is in Xu, Phoa, and Wong (2009), and a systematic construction method for nonregular fractional factorial designs of the required size is in Phoa and Xu (2009) and Phoa (2012).

The main contribution of the SRRS method is in its factor screening procedure. For most of the experiments or data, all-subset search is not feasible, so the factor screening procedure of SRRS aims at greatly reducing the number of insignificant factors. After factor dimension reduction, all-subset search may be feasible. If the dimension is still too large for the all-subset search to be feasible, there exists such efficient variable selection methods as the DS (Phoa, Pan, and Xu (2009)) that can be used. However, these efficient methods have already sacrificed accuracy when compared to all-subset search, in order to be efficient. The choice of these approaches for the model searching procedure of SRRS is unknown and needs further investigation.

The SRRS method can be easily modified and extended to the analysis of experiments other than SSDs. For examples, it can be extended for analyzing nonregular fractional factorial designs with interactions. Given factor sparsity and effect heredity assumptions, the calculations needed to carry out the analysis are easily performed with little computation time. It can also be extended for analyzing experiments with multiple responses.

Acknowledgement

This work was supported by National Science Council of Taiwan ROC grant number 100-2118-M-001-002-MY2. The author cordially thanks an associate editor, two referees, Dr. Hongquan Xu, and Dr. Jung-Shiang Hwang for their valuable discussions and suggestions.

References

Abraham, B., Chipman, H. and Vijayan, K. (1999). Some risks in the construction and analysis of supersaturated designs. *Technometrics* **41**, 135-141.

Beattie, S. D., Fong, D. K. H. and Lin, D. K. J. (2002). A two-stage Bayesian model selection strategy for supersaturated designs. *Technometrics* **44**, 55-63.

Booth, K. H. V. and Cox, D. R. (1962). Some systematic supersaturated designs. *Technometrics* **4**, 489-495.

Box, G. E. P. and Meyer, R. D. (1986). An analysis for unreplicated fractional factorials. *Technometrics* 28, 11-18.

Candes, E. J. and Tao, T. (2007). The Dantzig selector: statistical estimation when p is much larger than n. *Ann. Statist.* **35**, 2313-2351.

Chipman, H., Hamada, H. and Wu, C. F. J. (1997). A Bayesian variable-selection approach for analyzing designed experiments with complex aliasing. *Technometrics* **39**, 372-381.

Fan, J. and Lv, J. (2008). Sure independence screening for ultra-high dimensional feature space (with discussion). *J. Roy. Statist. Soc. Ser. B* **70**, 849-911.

George, E. I. and McMulloch, R. E. (1993). A statistical view of some chemometrics regression tools. *Technometrics* **35**, 109-148.

Georgiou, S. D. (2008). Modelling by supersaturated designs. *Comput. Statist. Data Anal.* **53**, 428-435.

Holcomb, D. R., Montgomery, D. C. and Carlyle, W. M. (2003). Analysis of supersaturated designs. *J. Quality Tech.* **35**, 13-27.

Hurvich, C. M. and Tsai, C.-L. (1989). Regression and time series model selection in small samples. *Biometrika* **76**, 297-307.

Hwang, J. S. and Hu, T. H. (2012). Stepwise paring down variation for identifying inflluential multi-factor interactions related to a continuous response variable. *Statistics in Biosciences* **4**, 197-212.

Li, R. and Lin, D. K. J. (2002). Data analysis in supersaturated designs. *Statist. Probab. Lett.* **59**, 135-144.

Li, R. and Lin, D. K. J. (2003). Analysis methods for supersaturated design: some comparisons. *J. Data Science* **1**, 249-260.

Lin, D. K. J. (1993). A new class of supersaturated designs. *Technometrics* **35**, 28-31.

Liu, Y. and Liu, M.Q. (2011). Construction of optimal supersaturated design with large number of levels. *J. Statist. Plann. Inference* **141**, 2035-2043.

Lu, X. and Wu, X. (2004). A strategy of searching active factors in supersaturated screening experiments. *J. Quality Tech.* **36**, 392-399.

Marley, C. J. and Woods, D. C. (2010). A comparison of design and model selection methods for supersaturated experiments. *Comput. Statist. Data Anal.* **54**, 3158-3167.

Phoa, F. K. H. (2012). A code arithmetic approach for Quaternary code designs and its application to $(1/64)^{th}$-fractions. *Ann. Statist.* **40**, 3161-3175.

Phoa, F. K. H., Pan, Y. H. and Xu, H. (2009). Analysis of supersaturated designs via the Dantzig selector. *J. Statist. Plann. Inference* **139**, 2362-2372.

Phoa, F. K. H. and Xu, H. (2009). Quater-fraction factorial design constructed via quaternary codes. *Ann. Statist.* **37**, 2561-2581.

Rais, F., Kamoun, A., Chaabouni, M., Claeys-Bruno, M., Phan-Tan-Luu, R. and Sergent, M. (2009). Supersaturated design for screening factors in sequencing the preparation of sulfated amides of olive pomace oil fatty acids. *Chemometr. Intell. Lab.* **99**, 71-78.

Sun, F., Lin, D. K. J., and Liu, M. Q. (2011). On construction of optimal mixed-level supersaturated designs. *Ann. Statist.* **39**, 1310-1333.

Satterthwaite, F. E. (1959). Random balance experimentation. *Technometrics* **1**, 111-137.

Westfall, P. H., Young, S. S., and Lin, D. K. J. (1998). Forward selection error control in the analysis of supersaturated designs. *Statist. Sinica* **8**, 101-117.

Williams, K. R. (1968). Designed experiments. *Rubber Age* **100**, 65-71.

Xu, H., Phoa, F. K. H. and Wong, W. K. (2009). Recent developments in nonregular fractional factorial designs. *Statist. Surveys* **3**, 18-46.

Zhang, Q. Z., Zhang, R. C. and Liu, M. Q. (2007). A method for screening active effects in supersaturated designs. *J. Statist. Plann. Inference* **137**, 235-248.

Institute of Statistical Science, Academia Sinica, Taipei City 115, Taiwan R.O.C.

E-mail: fredphoa@stat.sinica.edu.tw

(Received November 2011; accepted November 2012)