



**QUEEN'S
UNIVERSITY
BELFAST**

Mean squared error vs. frame potential for unsupervised variable selection

Zocco, F., & McLoone, S. (2017). Mean squared error vs. frame potential for unsupervised variable selection. In *Intelligent Computing, Networked Control, and Their Engineering Applications - International Conference on Life System Modeling and Simulation, LSMS 2017 and International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2017, Proceedings* (Vol. 762, pp. 353-362). (Communications in Computer and Information Science; Vol. 762). Springer Verlag. https://doi.org/10.1007/978-981-10-6373-2_36

Published in:

Intelligent Computing, Networked Control, and Their Engineering Applications - International Conference on Life System Modeling and Simulation, LSMS 2017 and International Conference on Intelligent Computing for Sustainable Energy and Environment, ICSEE 2017, Proceedings

Document Version:

Peer reviewed version

Queen's University Belfast - Research Portal:

[Link to publication record in Queen's University Belfast Research Portal](#)

Publisher rights

© 2017 Springer International Publishing AG.

This work is made available online in accordance with the publisher's policies. Please refer to any applicable terms of use of the publisher.

General rights

Copyright for the publications made accessible via the Queen's University Belfast Research Portal is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The Research Portal is Queen's institutional repository that provides access to Queen's research output. Every effort has been made to ensure that content in the Research Portal does not infringe any person's rights, or applicable UK laws. If you discover content in the Research Portal that you believe breaches copyright or violates any law, please contact openaccess@qub.ac.uk.

Open Access

This research has been made openly available by Queen's academics and its Open Research team. We would love to hear how access to this research benefits you. – Share your feedback with us: <http://go.qub.ac.uk/oa-feedback>

Mean Squared Error vs. Frame Potential for Unsupervised Variable Selection

Federico Zocco and Seán McLoone

School of Electronics, Electrical Engineering and Computer Science, Queen's University Belfast, Northern Ireland {fzocco01,s.mcloone}@qub.ac.uk

Abstract. Forward Selection Component Analysis (FSCA) provides a pragmatic solution to the NP-hard unsupervised variable selection problem, but is not guaranteed to be optimal due to the multi-modal nature of the mean squared error (MSE) selection metric used. Frame potential (FP) is a metric that has recently been shown to yield near-optimal greedy sensor selection performance for linear inverse problems. This paper explores if FP offers similar benefits in the unsupervised variable selection context. In addition, the backward elimination counterpart of FSCA is introduced for the first time (BECA) and compared with forward and backward FP based variable selection on a number of simulated and real world datasets. It is concluded that FP does not improve on FSCA and that while BECA yields comparable results to FSCA it is not a competitive alternative due to its much higher computational complexity.

Keywords: *Unsupervised dimensionality reduction, variable selection, frame potential, greedy algorithm*

1 Introduction

Principal Component Analysis (PCA) is a powerful technique for unsupervised dimensionality reduction, but as the resulting reduced representation is a linear combination of all variables, it is not, in general, straightforward to use it to identify a subset of key variables [2]. This is especially true if there is a high level of correlation among candidate variables. Various enhancements to PCA have been developed, such as SCoTLASS [4], DSPCA [5], sparse PCA [6]. These attempt to constrain PCA to produce sparse solutions, but are generally computationally intensive and do not necessarily yield good variable selection.

Directly solving the unsupervised variable selection problem is an NP-hard combinatorial optimization problem, and quickly becomes intractable as the number of candidate variables increases. Therefore, sub-optimal approaches which approximately solve the problem are generally employed. These fall into three categories: convex optimization relaxations of the cardinality constraint such as [7], heuristic approaches such as the aforementioned PCA enhancements, and greedy search methods such as Forward Selection Component Analysis (FSCA) [3].

Mathematically the unsupervised variable selection problem can be stated as follows. Given a dataset $\mathbf{X} \in \mathbb{R}^{m \times v}$ with m measurements of v variables, and an index set $\mathcal{N} = \{1, 2, \dots, v\}$ of the columns of \mathbf{X} we wish to find the subset \mathcal{S} of length k of the columns of \mathbf{X} that best approximates \mathbf{X} according to some metric, $G(\mathcal{S})$, that is:

$$\mathcal{S}^* = \underset{\mathcal{S} \subset \mathcal{N}, |\mathcal{S}|=k}{\operatorname{arg\,min}} G(\mathcal{S}) \quad (1)$$

In algorithms such as FSCA and sparse PCA $G(\mathcal{S})$ is the reconstruction error of the optimum linear reconstruction of \mathbf{X} ($= \mathbf{X}([\mathcal{N}]$) by $\mathbf{X}([\mathcal{S}]$), that is

$$G(\mathcal{S}) = \frac{1}{mv} \|\mathbf{X} - \hat{\mathbf{X}}([\mathcal{S}])\|_F^2 \quad (2)$$

where $\hat{\mathbf{X}}([\mathcal{S}]) = \mathbf{X}([\mathcal{S}]\mathbf{X}([\mathcal{S}])^\dagger \mathbf{X}$. Here, for a generic matrix \mathbf{M} , the notation \mathbf{M}^\dagger signifies the Moore-Penrose pseudoinverse of \mathbf{M} , $\mathbf{M}([\mathcal{R}])$ is a matrix containing only the columns of \mathbf{M} specified by index set \mathcal{R} , and $\|\mathbf{M}\|_F$ is the Frobenius norm of \mathbf{M} . Essentially (2) is the mean squared error (MSE) of the approximation, or equivalently a proxy for the variance explained metric $V_{\mathbf{X}}(\mathcal{S})$ if the columns of \mathbf{X} are mean-centred, that is:

$$V_{\mathbf{X}}(\mathcal{S}) = 100 \cdot \left(1 - mv \frac{G(\mathcal{S})}{\|\mathbf{X}\|_F^2} \right) \quad (3)$$

Hence, minimizing $G(\mathcal{S})$ is equivalent to maximizing the variance explained.

Recently a metric referred to as the frame potential (FP) has been shown to guarantee near-optimal sensor selection solutions in terms of MSE for linear inverse problems when using a greedy selection algorithm [8]. The frame potential of a matrix \mathbf{X} is defined as

$$FP(\mathbf{X}) = \sum_{i,j=1}^v |\langle \mathbf{x}_i, \mathbf{x}_j \rangle|^2 = \sum_{i,j=1}^m |\langle \vec{\mathbf{x}}_i, \vec{\mathbf{x}}_j \rangle|^2 \quad (4)$$

where \mathbf{x}_i and $\vec{\mathbf{x}}_i$ are the i -th column and i -th row of \mathbf{X} , respectively [12], and minimizing it encourages orthogonality among the selected variables. Its attractiveness with regard to greedy selection algorithms arises from the *submodularity* property [8], [9], [10], which is defined as follows.

Definition 1. (*Submodular function [8]*) Given three sets \mathcal{X} , \mathcal{Y} and \mathcal{N} such that $\mathcal{X} \subset \mathcal{Y} \subset \mathcal{N}$ and given an element $i \in \mathcal{N} \setminus \mathcal{Y}$, a function G is submodular if it satisfies $G(\mathcal{X} \cup i) - G(\mathcal{X}) \geq G(\mathcal{Y} \cup i) - G(\mathcal{Y})$.

The significance of this property is that in the field of combinatorial optimization it has been proven that greedy maximization of normalised, monotone, submodular functions is near-optimal in the sense of being bounded to be within a factor $(1 - e^{-1})$ of the global maximum [11], that is:

$$G(\mathcal{S}_{\text{greedy}}) \geq (1 - e^{-1})G(\mathcal{S}_{\text{opt}}) \quad (5)$$

where $\mathcal{S}_{\text{opt}} = \arg \max_{\mathcal{S} \subset \mathcal{N}, |\mathcal{S}|=k} G(\mathcal{S})$. Ranieri et al. [8] showed that the FP based cost function

$$F(\mathcal{Q}) = FP(\mathbf{X}) - FP(\mathbf{X}([\mathcal{N} \setminus \mathcal{Q}])) \quad (6)$$

is a normalized, monotone and submodular function with respect to \mathcal{Q} , and hence satisfies the optimality bound requirements. Maximizing $F(\mathcal{Q})$ is equivalent to minimizing $FP(\mathcal{S})$, where $\mathcal{S} = \mathcal{N} \setminus \mathcal{Q}$.

The sensor selection problem considered by Ranieri et al. [8] can be expressed as $\mathbf{f} = \mathbf{X}\boldsymbol{\alpha}$, where $\mathbf{f} \in \mathbb{R}^{m \times 1}$ is the measured physical field, $\boldsymbol{\alpha} \in \mathbb{R}^{v \times 1}$ are the parameters to be estimated and $\mathbf{X} \in \mathbb{R}^{m \times v}$ is a known linear model. Then, given a limited number of sensor locations $r < m$ corresponding to the rows specified in the index set $\vec{\mathcal{S}}$, and denoting $\mathbf{X}([\vec{\mathcal{S}}])$ as the matrix containing the rows of \mathbf{X} indexed by $\vec{\mathcal{S}}$, solving the sensor placement problem equates to minimizing

$$G(\vec{\mathcal{S}}) = MSE(\hat{\boldsymbol{\alpha}}(\vec{\mathcal{S}})) = \|\hat{\boldsymbol{\alpha}}(\vec{\mathcal{S}}) - \boldsymbol{\alpha}\|_2^2 \quad (7)$$

with respect to $\vec{\mathcal{S}}$, subject to $|\vec{\mathcal{S}}| = r$, where $\hat{\boldsymbol{\alpha}}(\vec{\mathcal{S}}) = \mathbf{X}([\vec{\mathcal{S}}])^\dagger \mathbf{f}([\vec{\mathcal{S}}])$.

Ranieri et al. [8] showed that by employing a backward elimination greedy search algorithm to maximize (6) near optimal MSE performance can be obtained. It should be noted that a near-optimal solution in terms of FP does not guarantee near-optimality with regards to the MSE. In the case of the sensor selection inverse problem considered in [8], the authors show that a strong link exists between the MSE of the model parameters being estimates ($\boldsymbol{\alpha}$) and FP, with the result that a near-optimal solution is guaranteed for that specific problem formulation. However, for the general unsupervised variable selection problem, where the focus is on the MSE of the reconstruction of \mathbf{X} , and not the MSE of the model parameters, such links are not readily established and hence MSE performance is not guaranteed. Nevertheless, it is interesting to explore if optimizing with respect to FP with its *orthogonality encouraging* property provides performance benefits over FSCA, which is a greedy forward selection algorithm that directly optimizes with respect to the reconstruction MSE. As such, this paper presents an empirical investigation of the performance of FP versus reconstruction MSE as a variable selection metric for unsupervised variable selection problems. The standard FSCA algorithm performs variable selection using forward selection, while the FP algorithm in [8] uses backward elimination. Therefore, to enable a fair comparison, forward and backward versions of each algorithm are considered. These are denoted FSCA, BECA, FSFP and BEFP.

The remainder of the paper is organized as follows: Section 2 introduces the greedy selection algorithms under investigation. Section 3 presents the results of two simulated experiments, while Sect. 4 presents the results for two real world cases. Finally, conclusions are provided in Sect. 5. The nomenclature used is as follows: matrices are denoted by bold capital letters, vectors are denoted by bold lowercase letters, \mathbf{X}_0 and \mathbf{X}_u are the data matrix variants of \mathbf{X} with zero mean columns and one norm columns, respectively.

2 Algorithms

The four greedy unsupervised variable selection algorithms investigated are:

1. Forward Selection Component Analysis (**FSCA**): The implementation used here is equivalent to, but less computationally efficient than the algorithms presented in [3], and is used to allow direct comparison with the FP selection criteria.
2. Backward Elimination Component Analysis (**BECA**): This implements the classic backward elimination algorithm, where all variables are initially selected and then the least important variables are successively eliminated using the change in reconstruction MSE as the selection metric.
3. Forward Selection Frame Potential (**FSFP**): This implements the forward selection approach using $F(\mathcal{Q})$ in (6) as the selection metric.
4. Backward Elimination Frame Potential (**BEFP**): This uses the backward elimination approach of BECA but with MSE replaced by $F(\mathcal{Q})$ as the selection metric.

The pseudo-code for the forward selection and backward elimination algorithms is presented in General algorithm 1 and General algorithm 2, respectively, where the *Selector* flag allows selection of the specific algorithm of interest. Note that the FP algorithms use \mathbf{X}_u instead of \mathbf{X} , as recommended in [8], as this has been found to improve selection performance.

General algorithm 1 Forward greedy variable selection

Input: $\mathbf{X}, k, Selector$

Output: \mathbf{s}

Initialisation: $\mathbf{s} = [], \mathbf{a} = [1, 2, 3, \dots, v]$

for $j = 1$ to k **do**

switch *Selector* **do**

case FSCA

(a) $\hat{\Theta}(a_i) = \mathbf{X}_0([\mathbf{s}, a_i])^\dagger \mathbf{X}_0$

(b) $\hat{\mathbf{X}}_0(a_i) = \mathbf{X}_0([\mathbf{s}, a_i])\hat{\Theta}(a_i)$

(c) $a_{i^*} = \arg \min_{a_i} \|\mathbf{X}_0 - \hat{\mathbf{X}}_0(a_i)\|_F^2$

case FSFP

(a) $a_{i^*} = \arg \max_{a_i} \{FP(\mathbf{X}_u) - FP(\mathbf{X}_u([\mathbf{s}, a_i]))\}$

end switch

$\mathbf{s} = [\mathbf{s}, a_{i^*}], \mathbf{a} = \mathbf{a} \setminus a_{i^*}$

end for

In all experiments conducted, algorithm performance is evaluated using the percentage of dataset variance explained by reconstruction using the selected variables, as defined in (3) (with $\mathbf{X} = \mathbf{X}_0$).

General algorithm 2 Backward greedy variable elimination

Input: $\mathbf{X}, k, Selector$

Output: \mathbf{s}

```

Initialisation:  $\mathbf{s} = [], \mathbf{a} = [1, 2, 3, \dots, v]$ 
for  $j = 1$  to  $(v - k)$  do
  switch  $Selector$  do
    case BECA
      (a)  $\hat{\Theta}(a_i) = \mathbf{X}_0([\mathbf{a} \setminus a_i])^\dagger \mathbf{X}_0$ 
      (b)  $\hat{\mathbf{X}}_0(a_i) = \mathbf{X}_0([\mathbf{a} \setminus a_i])\hat{\Theta}(a_i)$ 
      (c)  $a_{i^*} = \arg \min_{a_i} \|\mathbf{X}_0 - \hat{\mathbf{X}}_0(a_i)\|_F^2$ 
    case BEFP
      (a)  $a_{i^*} = \arg \max_{a_i} \{FP(\mathbf{X}_u) - FP(\mathbf{X}_u([\mathbf{a} \setminus a_i]))\}$ 
  end switch
   $\mathbf{a} = \mathbf{a} \setminus a_{i^*}$ 
end for
 $\mathbf{s} = \mathbf{a}$ 

```

3 Simulated Case Studies

This section presents results comparing FSCA, BECA, FSFP and BEFP on two simulated case studies introduced in [3]. Since PCA provides an upper bound on the achievable variance explained by variable selection algorithms for a fixed number of components [3], the algorithms are also benchmarked against PCA. Following [1], the performance comparison focuses on variable selection to achieve 99% variance explained. The number of variables/components needed to achieve 99% variance explained is denoted as $k_{99\%}$.

3.1 Simulated Dataset 1: Four Distinct Variables

The generation of this dataset begins by defining four base variables $w_0, x_0, y_0, z_0 \sim N(0, 1)$, 20 noise variables $\epsilon_1, \dots, \epsilon_{20} \sim N(0, 0.1)$ and two larger noise variables $\epsilon_{21}, \epsilon_{22} \sim N(0, 0.4)$. These are then used to define the dataset $\mathbf{X} = [\mathbf{w}_0, \dots, \mathbf{w}_5, \mathbf{x}_0, \dots, \mathbf{x}_5, \mathbf{y}_0, \dots, \mathbf{y}_5, \mathbf{z}_0, \dots, \mathbf{z}_5, \mathbf{h}_1, \mathbf{h}_2] \in \mathbb{R}^{n \times 26}$, where the columns of \mathbf{X} are n realizations of the variables $\{w_i = w_0 + \epsilon_i\}_{i=1, \dots, 5}$, $\{x_i = x_0 + \epsilon_{i+5}\}_{i=1, \dots, 5}$, $\{y_i = y_0 + \epsilon_{i+10}\}_{i=1, \dots, 5}$, $\{z_i = z_0 + \epsilon_{i+15}\}_{i=1, \dots, 5}$, $h_1 = w_0 + x_0 + \epsilon_{21}$ and $h_2 = y_0 + z_0 + \epsilon_{22}$. Hence, the dataset is highly redundant with just 4 independent variables. The algorithms have been evaluated over 200 realizations of \mathbf{X} with $n = 1000$. The results reported are the average performance over the 200 realizations.

Figure 1(a) shows the percentage variance explained (V_X) with respect to the variance explained by PCA (V_{PCA}) as a function of k . Among the four algorithms, only FSCA shows a significantly different trend, rapidly achieving 95%, but then showing no improvement until $k = 5$. With reference to the $k_{99\%}$ results in Table 1, FSCA performs best, followed by BECA and FSFP. BEFP is inferior to the other algorithms requiring 21 variables to achieve the 99% target.

3.2 Simulated dataset 2: Block Redundancy

This dataset is defined as follows: $\mathbf{X} = [\mathbf{X}^0, \mathbf{X}^1] \in \mathbb{R}^{m \times v}$, where $\mathbf{X}^0 \in \mathbb{R}^{m \times u}$ with $\mathbf{X}_{i,j}^0 \sim N(0, 1)$, $\mathbf{X}^1 = \mathbf{X}^0 \cdot \Phi + \mathbf{E}$, $\Phi \in \mathbb{R}^{u \times (v-u)}$ with $\Phi_{i,j} \sim N(0, 1)$ and $\mathbf{E} \in \mathbb{R}^{m \times (v-u)}$ with $\mathbf{E}_{i,j} \sim N(0, 0.1)$. Hence, this dataset contains $u \leq v$ independent variables. Three different (u, v) combinations are considered: (10, 30), (15, 50) and (20, 75). Results are presented in each case for the average of 200 realizations of \mathbf{X} with $m = 1000$.

Figure 1(b) shows the performance of each algorithm relative to PCA for different values of k when $u = 10$. As can be seen, FSCA yields the best results followed closely by BECA. Both FP algorithms perform poorly for this problem when $k < 10$, with the forward implementation superior to the backward implementation. As k increases the differences between the algorithms decrease, and they all converge at $k = 10$. Table 2 shows the $k_{99\%}$ results and the corresponding V_X in parentheses for the different (u, v) combinations. All algorithms need u variables to achieve the 99% threshold. The difference between the worst and the best is less than 0.2%. To demonstrate the robustness of the results, box plots showing the distribution of the variance explained by each algorithm over the 200 realizations are presented in Fig. 2 for the case $u = 10$, for $k = 5$ and $k = 10$.

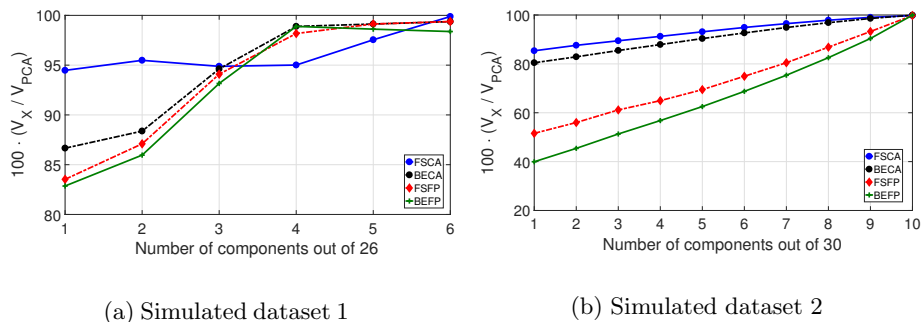


Fig. 1: Percentage explained variance with respect to V_{PCA} for different values of k for: (a) Simulated dataset 1 and; (b) Simulated dataset 2 for $u = 10$.

Table 1: (Simulated dataset 1) Values of $k_{99\%}$ achieved and the corresponding V_X . In bold the winner algorithm after PCA.

	PCA	FSCA	BECA	FSFP	BEFP
$k_{99\%}$	5	6	8	8	21
V_X	99.01	99.30	99.07	99.06	99.14

Table 2: (Simulated dataset 2) Values of $k_{99\%}$ achieved and, between brackets, the corresponding V_X .

u	v	PCA	FSCA	BECA	FSFP	BEFP
10	30	10(99.95)	10(99.72)	10(99.86)	10(99.75)	10(99.89)
15	50	15(99.96)	15(99.76)	15(99.89)	15(99.69)	15(99.93)
20	75	20(99.97)	20(99.77)	20(99.90)	20(99.68)	20(99.95)

4 Real World Case Studies

In this section the algorithms are compared on two different real world datasets: wafer metrology data from a semiconductor manufacturing process and gas sensor array data collected from a gas delivery system.

4.1 Real Dataset 1: Wafer Site Optimization

This dataset, which is described in detail in [1], consists of wafer metrology data from a semiconductor manufacturing process for a set of 316 wafers. The goal is to improve the efficiency of wafer profile monitoring by reducing the number of measured sites from a candidate set of 50 sites, without discarding valuable information. This can be cast as a variable selection problem, with the columns of the data matrix representing the candidate sites and the rows representing the measurements from individual wafers, hence $\mathbf{X} \in \mathbb{R}^{316 \times 50}$. The goal is to select the number of sites needed to achieve 99% variance explained. The results of the analysis of this dataset are presented in Fig. 3(a) and in Tables 3 and 4.

Figure 3(a) and Table 3 show that the MSE-based algorithms (FSCA and BECA) substantially outperform the FP-based algorithms for this problem achieving consistently higher explained variance for a given number of selected variables. FSCA yields the best performance for low values of k while BECA is marginally superior at higher values. Both algorithms identify 7 sites as necessary to achieve 99% variance explained. In contrast, the best FP-based result is 10 sites.

As described in [1], the unmeasured sites can be estimated from the measured sites using a linear model with the selected sites as regressors. Table 4 provides a comparison of the performance of the seven sites selected by FSCA and BECA with regard to their ability to predict the 43 unmeasured sites on each wafer. Performance is expressed in terms of the mean, standard deviation and minimum variance explained over the 43 unmeasured sites. As can be seen, BECA selects completely different sites to FSCA, and in this instance produces significantly better results.

4.2 Real Dataset 2: Gas Sensor Array Analysis

This dataset was selected from the UCI Machine Learning Repository [13], [14] and consists of measurements from 16 chemical sensors exposed to 6 gases at different concentration levels gathered in a gas delivery platform facility at the BioCircuits Institute, University of California, San Diego. From each sensor, 8 features have been extracted giving a total of 128 variables plus the concentration of the analyte gas. Here, only the data from batch 3 is considered giving a dataset $\mathbf{X} \in \mathbb{R}^{1586 \times 129}$ for analysis. Due to the large variation in the magnitude of the variables in the dataset, all variables were standardized prior to analysis. Results are presented in Fig. 3(b) and Table 5 for the standardized dataset.

Here again, the MSE-based algorithms outperform FP, particularly for lower numbers of selected variables. In addition, to achieve 99% variance explained FSFP and BEFP need 17 variables while FSFA and BECA achieve the same performance with only 12 variables.

Table 3: (Real dataset 1) Values of $k_{99\%}$ achieved and the corresponding V_X .

	PCA	FSCA	BECA	FSFP	BEFP
$k_{99\%}$	5	7	7	10	11
V_X	99.07	99.02	99.17	99.24	99.17

Table 4: (Real dataset 1) Mean, minimum and standard deviation of V_X for the unmeasured sites using the sites selected by FSCA and BECA as regressors.

	Best 7 sites	Mean V_X	Min V_X	σ
FSCA	45, 27, 1, 24, 9, 49, 14	98.74	93.74	1.34
BECA	4, 15, 16, 31, 42, 44, 48	99.06	96.70	0.86

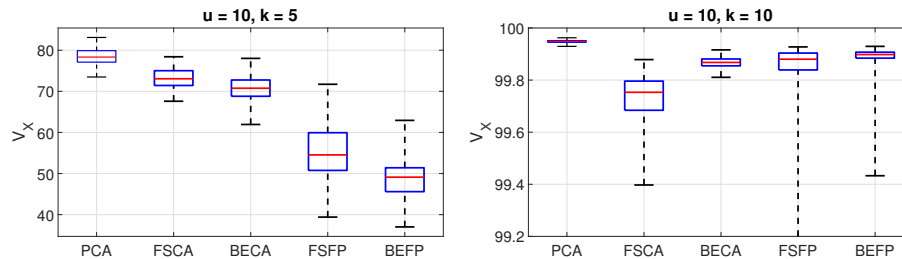


Fig. 2: (Simulated dataset 2) Boxplot of V_X by each algorithm for the case $u = 10$ with $k = 5$ (left) and $k = 10$ (right).

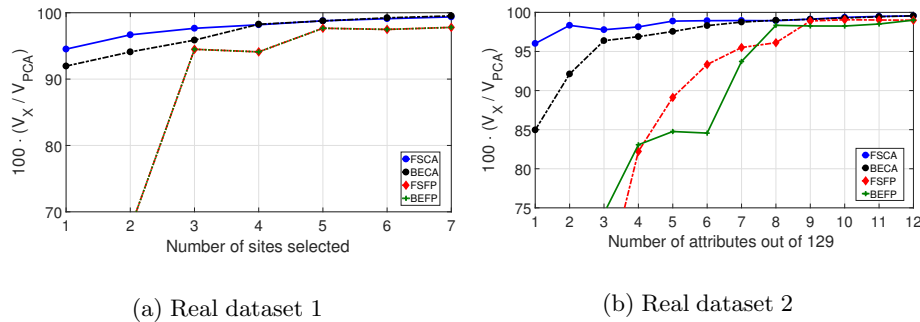


Fig. 3: Percentage explained variance with respect to V_{PCA} for different values of k for: (a) Real dataset 1 (left) and; (b) Real dataset 2 (right).

Table 5: (Real dataset 2) Values of $k_{99\%}$ achieved and the corresponding V_X .

	PCA	FSCA	BECA	FSFP	BEFP
$k_{99\%}$	8	12	12	17	17
V_X	99.00	99.05	99.07	99.02	99.01

5 Conclusions

Motivated by its effectiveness in a recently reported optimum sensor placement application, this paper has explored the value of FP as an alternative to MSE as a selection metric in unsupervised variable selection problems. Both forward selection and backward elimination greedy selection algorithms have been considered. The results of two simulated and two real case studies show that, in general, for a fixed number of selected variables, selection based on MSE significantly outperforms FP, and therefore FP is not a good choice for this type of problem. An explanation for this is that while greedy FP-based algorithms enjoy guarantees on near-optimal solutions in terms of FP this does not automatically extend to guarantees on MSE performance. In general FSCA outperforms BECA when the number of variables selected is low, but BECA has comparable and sometimes marginally better performance when approaching 99% variance explained. However, FSCA has the advantage of being computationally much more efficient, and therefore is the preferred algorithm.

Acknowledgments

The first author gratefully acknowledges Irish Manufacturing Research (IMR) for its financial support of his PhD.

References

1. P. Prakash *et al.*, “Optimal wafer site selection using forward selection component analysis,” in *Advanced Semiconductor Manufacturing Conference (ASMC), 2012 23rd Annual SEMI*. IEEE, pp. 91–96, 2012.
2. G. P. McCabe, “Principal variables,” *Technometrics*, vol. 26, no. 2, pp. 137–144, 1984.
3. L. Puggini and S. McLoone, “Forward selection component analysis: algorithms and applications,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, DOI: 10.1109/TPAMI.2017.2648792, 2017.
4. I. T. Jolliffe *et al.*, “A modified principal component technique based on the lasso,” *Journal of Computational and Graphical Statistics*, vol. 12, no. 3, pp. 531–547, 2003.
5. A. d’Aspremont *et al.*, “A direct formulation for sparse pca using semidefinite programming,” *SIAM review*, vol. 49, no. 3, pp. 434–448, 2007.
6. H. Zou *et al.*, “Sparse principal component analysis,” *Journal of Computational and Graphical Statistics*, vol. 15, no. 2, pp. 265–286, 2006.
7. M. Masaeli *et al.*, “Convex principal feature selection,” *SDM. SIAM*, pp. 619–628, 2010.
8. J. Ranieri *et al.*, “Near-optimal sensor placement for linear inverse problems,” *Signal Processing, IEEE Transactions on*, vol. 62, no. 5, pp. 1135–1146, 2014.
9. A. Das and D. Kempe, “Algorithms for subset selection in linear regression,” in *Proc. ACM Symp. Theory Comput. (STOC)*, 2009.
10. A. Das and D. Kempe, “Submodular meets spectral: greedy algorithms for subset selection, sparse approximation and dictionary selection,” in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2011.
11. G. Nemhauser *et al.*, “An analysis of approximations for maximizing submodular set functions–I,” *Math. Prog.*, vol. 14, pp. 265–294, 1978.
12. S. Waldron, “Generalised welch bound equality sequences are tight frames,” Dept. Math., Univ. Auckland, Auckland, New Zealand, Tech. Rep., 2003.
13. A. Vergara *et al.*, “Chemical gas sensor drift compensation using classifier ensembles,” *Sensors and Actuators B: Chemical*, vol. 166, pp. 320–329, 2012.
14. I. Rodriguez-Lujan *et al.*, “On the calibration of sensor arrays for pattern recognition using the minimal number of experiments,” *Chemometrics and Intelligent Laboratory Systems*, vol. 130, pp. 123–134, 2014.