Supplementary Material of "Hierarchical Sparse Functional Principal Component Analysis for Multistage Multivariate Profile Data"

S.1 Determination of the Number of PCs

This section provides more information on how to select the number of PCs K discussed in Section 2.5. As our HSMFPCA is performed sequentially (see Figure 3), after updating $\mathbf{X}$ by $\mathbf{X}-\mathbf{X}\widehat{\mathbf{v}}\widehat{\mathbf{v}}^T$, a hypothesis test based on the Tracy-Widom convergence is used to check if the covariance matrix of the updated data $\mathbf{x}_i$ is an identity matrix. If rejected, we believe more signals exist in the new data matrix $\mathbf{X}$ and continue looking for the next PC. Specifically, if $\mathbf{x}_i \overset{\text{i.i.d}}{\sim} \mathrm{N}(0,\mathbf{I})$, $N_1=\max(N,P)-1$, $P_1=\min(N,P)$ and e_1 is the largest eigenvalue of $\mathbf{X}^T\mathbf{X}$, then we have the following theorem (El Karoui 2003, Akemann *et al.* 2011):

Theorem S.1. *If $N,P\to\infty$, and $P/N\to\rho\in(0,\infty)$, then $(e_1-\mu_{NP})/\sigma_{NP}\xrightarrow{d} TW_1$, where $\mu_{NP}=(\sqrt{N_1}+\sqrt{P_1})^2$, $\sigma_{NP}=(\sqrt{N_1}+\sqrt{P_1})(1/\sqrt{N_1}+1/\sqrt{P_1})^{1/2}$ and TW_1 is the Tracy-Widom distribution.*

We take the single factor model in Section 3.1 for example where $L=1$, $\sigma_\varepsilon^2=1$ and $\mathbf{v}_1$ is the eigenvector in Case 1 (see Figure 4). The results for the eigenvectors in Case 2-7 (see Figure 6 and 8) are also very similar. The signal size $\sigma_1\in\{1.0,1.1,\dots,5.0\}$ and the sample size $N\in\{20,50,100,200\}$. For each combination of σ_1 and N, $R=500$ random data sets are simulated, and for each data set, our proposed HSMFPCA is applied repeatedly as long as $e_1/\widehat{\sigma}_\varepsilon^2>Q_{TW_1}(0.95)\times\sigma_{NP}+\mu_{NP}$, where $\widehat{\sigma}_\varepsilon^2=\mathrm{Median}\big(\{\sum_{i=1}^N(\mathbf{x}_{i,sjt}-\boldsymbol{\mu}_{sjt})^2/N\}_{s,j,t}\big)$ (Ma 2013) and $Q_{TW_1}(0.95)$ is the 95% quantile of the Tracy-Widom distribution. The correct rate of the selected number of PCs in these simulated data sets is defined as CR $= \#(K = L)/R$, which is plotted in Figure S.1. The CR gets higher as the sample size or signal size increases. Even though $N \leq P$, it can be seen in Figure S.1 that the true number of PCs can still be correctly inferred at a high probability whenever the signal size is significantly larger than the noise size.

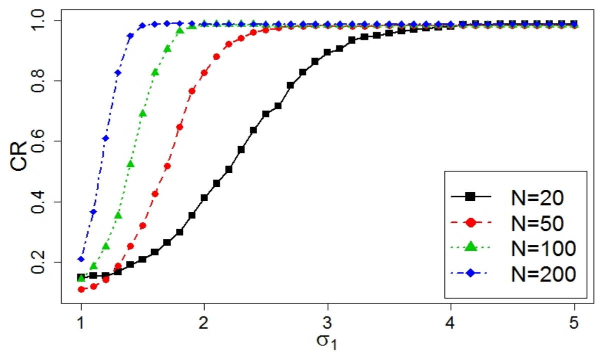

Figure S.1: Correct rates of the selected number of PCs for different sample and signal sizes.

S.2 Selection of the Tuning Parameters

We now verify that selecting the tuning parameters by the AIC(λ_3) in Section 2.5 is comparable to that by the AIC($\lambda_1, \lambda_2, \lambda_3$), taking the single factor model in Section 3.1 again for example. In our fast AIC(λ_3), we set $\lambda_3 \in \{0.0, 0.1, \ldots, 10.0\}$ and for each λ_3, $\lambda_2 = 10\lambda_3$, $\lambda_1 = 50\lambda_3$, while in the exact AIC($\lambda_1, \lambda_2, \lambda_3$), we let $\lambda_3 \in \{0.0, 0.1, \ldots, 10.0\}$, $\lambda_2 \in \{0, 1, \ldots, 100\}$, $\lambda_1 \in \{0, 5, \ldots, 500\}$. For one simulated date set, the best AIC is searched in the two configured grids above, the results of which are shown in Table S.1. It can be clearly seen that the optimal values of the AIC(λ_3) and the AIC($\lambda_1, \lambda_2, \lambda_3$) are approximately the same. Actually, the minimum value of AIC(λ_3) is smaller than the 99.84% of the grid values of AIC($\lambda_1, \lambda_2, \lambda_3$) in our simulations. Additionally, the selected tuning parameters λ_1, λ_2, λ_3 and the performance in terms of the criteria in Table 1 are also very similar. The estimated eigenvectors with the optimal tuning parameters are displayed in Figure S.2, where there is no obvious visual difference between the results of the AIC(λ_3) and the AIC($\lambda_1, \lambda_2, \lambda_3$).

Table S.1: Comparison results of the AIC(λ_3) and the AIC($\lambda_1, \lambda_2, \lambda_3$).

	Minimum value	*df*	(λ_1 λ_2 λ_3)	ZM	F1	Angle (10^{-2})	RMSE (10^{-3})	EV
AIC(λ_3)	10009.54	31	(340, 68, 6.8)	0.9650	0.8772	7.0312	7.8058	0.2563
AIC($\lambda_1, \lambda_2, \lambda_3$)	10007.91	31	(370, 88, 6.8)	0.9650	0.8772	7.0307	7.8052	0.2563

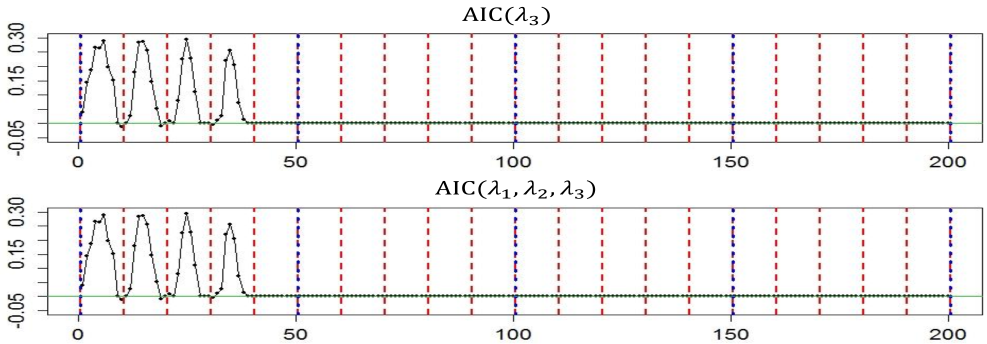

Figure S.2: Estimated eigenvectors by the AIC(λ_3) and the AIC($\lambda_1, \lambda_2, \lambda_3$).

S.3 HSMFPCA Performance for Different Sample and Signal Sizes

We evaluate the performance of our proposed HSMFPCA in terms of the criteria in Table 1 as the sample size N and the signal size σ_1 vary in Case 1-7 (see Figure 4, 6 and 8). Table S.2 and Figure S.3 show the results in Case 2, where it can be seen that as N and σ_1 increase, ZM and F1 get higher, indicating that the sparsity pattern of the eigenvector is more correctly identified, and Angle and RMSE become lower, implying that the element values of the eigenvector is more accurately estimated. For each value of σ_1, the variances of the results, shown as the lengths of the boxplots in Figure S.3, are also reduced when N gets larger. The results of the other cases are very similar, which are thus not shown here for saving space.

Table S.2: Performance of the HSMFPCA for different sample and signal sizes in Case 2.

	N	$\sigma_1 = 2$	$\sigma_1 = 5$	$\sigma_1 = 8$	$\sigma_1 = 10$	N	$\sigma_1 = 2$	$\sigma_1 = 5$	$\sigma_1 = 8$	$\sigma_1 = 10$
ZM		0.7617	0.9300	0.9420	0.9498		0.8458	0.9467	0.9554	0.9612
F1		0.4994	0.8322	0.8564	0.8737		0.6850	0.8659	0.8858	0.8993
Angle	20	0.7107	0.2247	0.1349	0.1071	50	0.4272	0.1348	0.0826	0.0653
RMSE		0.0707	0.0248	0.0150	0.0119		0.0461	0.0148	0.0092	0.0072
EV		0.0525	0.1292	0.2526	0.3400		0.0358	0.1206	0.2492	0.3384
ZM		0.8848	0.9440	0.9687	0.9782		0.8913	0.9590	0.9783	0.9834
F1		0.7490	0.8604	0.9172	0.9414		0.7591	0.8951	0.9424	0.9555
Angle	100	0.2951	0.0959	0.0560	0.0436	200	0.2102	0.0665	0.0394	0.0310
RMSE		0.0325	0.0106	0.0062	0.0048		0.0232	0.0074	0.0044	0.0034
EV		0.0295	0.1165	0.2446	0.3338		0.0267	0.1157	0.2449	0.3346

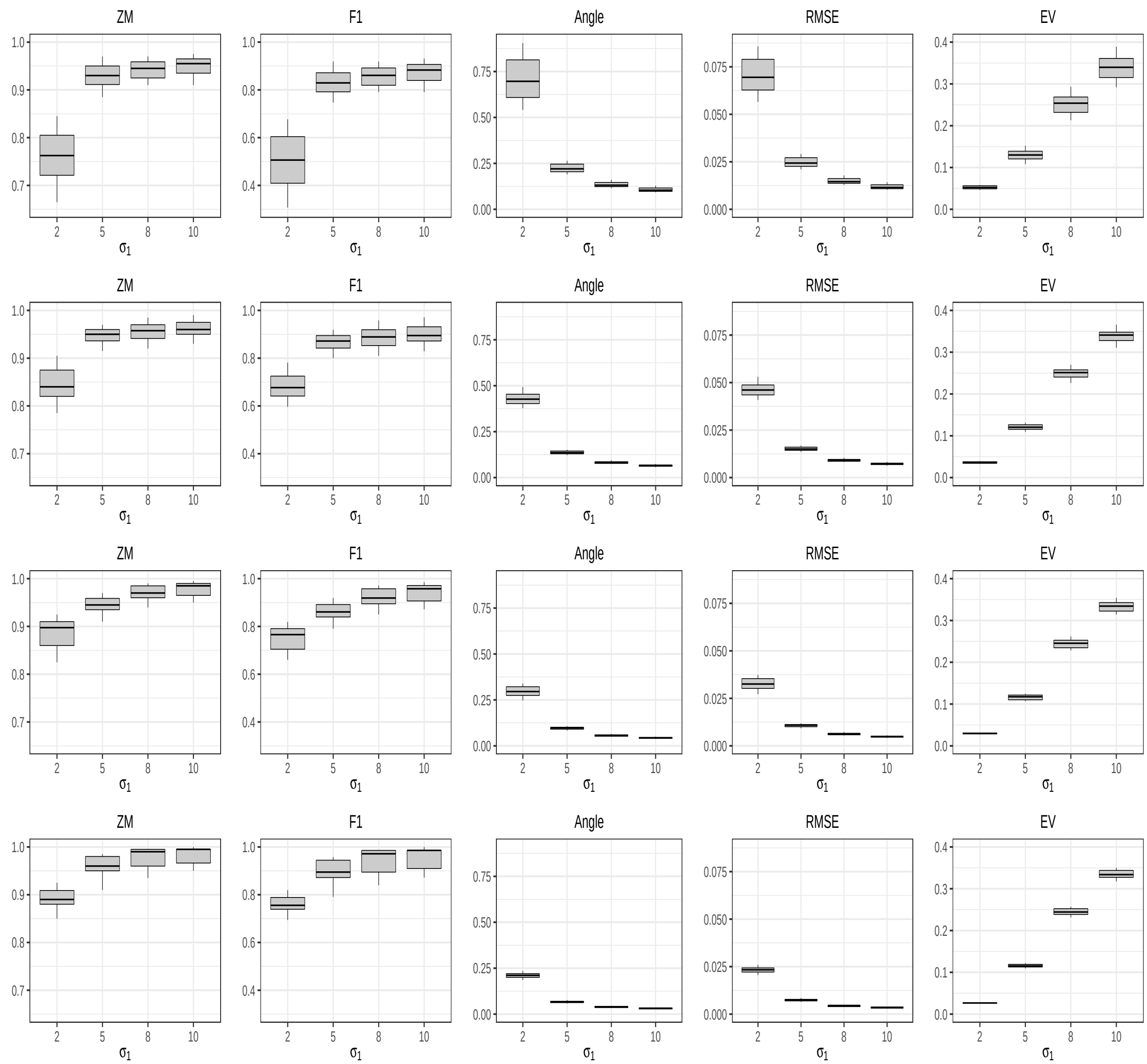

Figure S.3: Boxplots of the HSMFPCA Performance for different sample and signal sizes in Case 2. Row 1-4 correspond to $N = 20, 50, 100, 200$.

S.4 Real Example Study

The 120 samples of the original profile data of each process variable from the three stages in our real example (see Section 4) are plotted individually in Figure S.4. Please note that that different process variables are recorded by different measurement units and scales and they are also plotted by different ranges in the vertical axis in Figure S.4, so the observations of the data variance directly from Figure S.4 is very misleading. We follow the guidelines in Section 2.5 to perform data preprocessing, and the centered and standardized profile data have been shown in Figure 10.

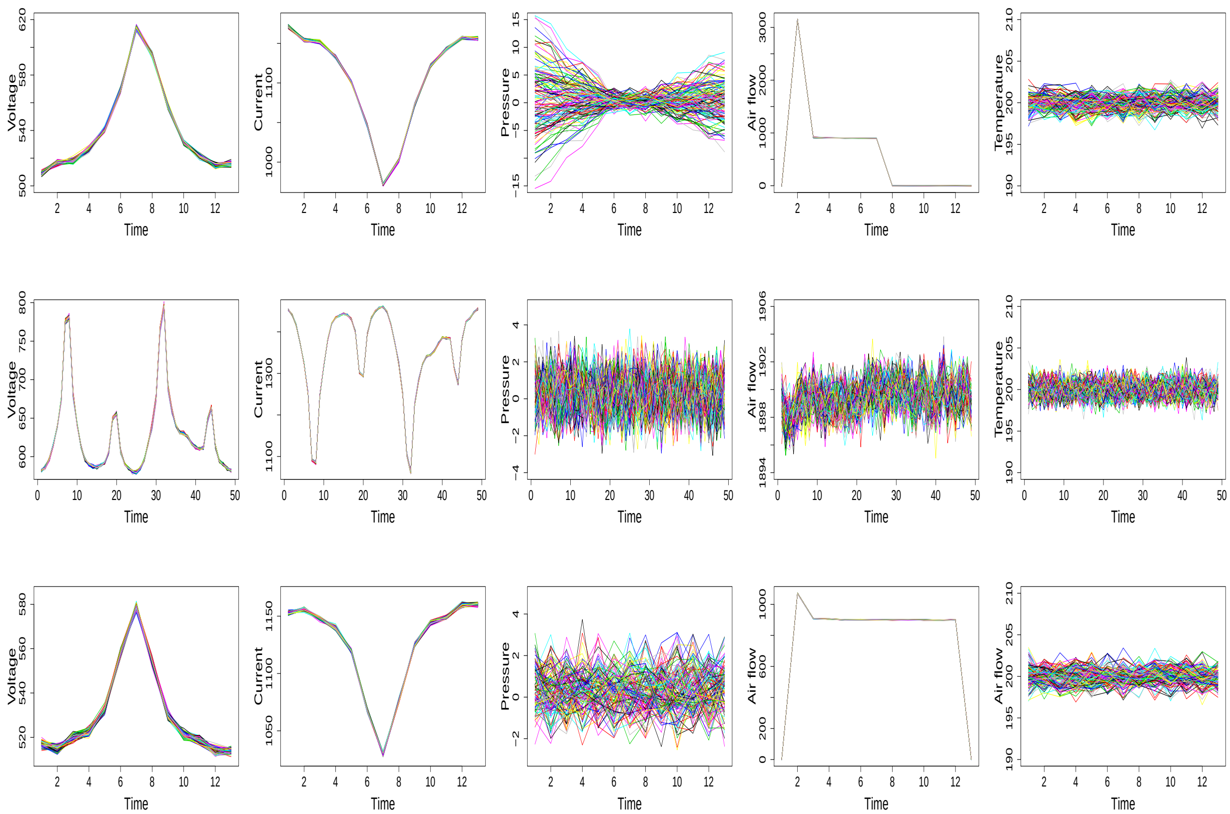

Figure S.4: Profile data of process variables from all stages. Row 1-3 correspond to Stage 1-3.

We also apply the VPCA, SSPCA, ESMFPCA and PSMFPCA to the real example data set, the results of which are plotted in Figure S.5-S.8, respectively. Compared to these competing methods, though the percents of explained variance are reduced a little bit, our HSMFPCA in Figure 11 is shown to be much better in generating much more sparse and more interpretable eigenvectors.

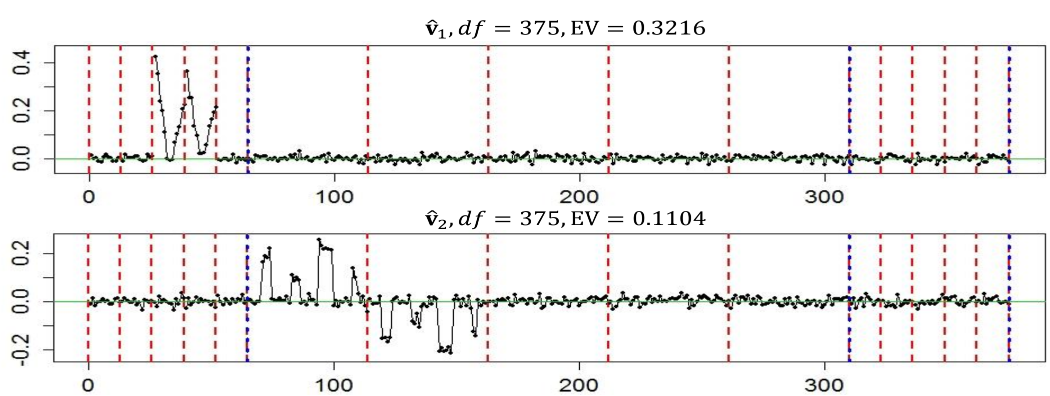

Figure S.5: Estimated eigenvectors by the VPCA.

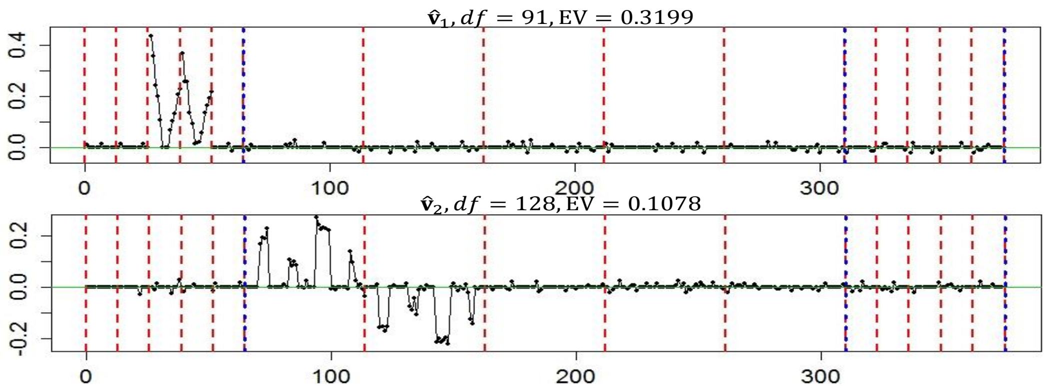

Figure S.6: Estimated eigenvectors by the SSPCA.

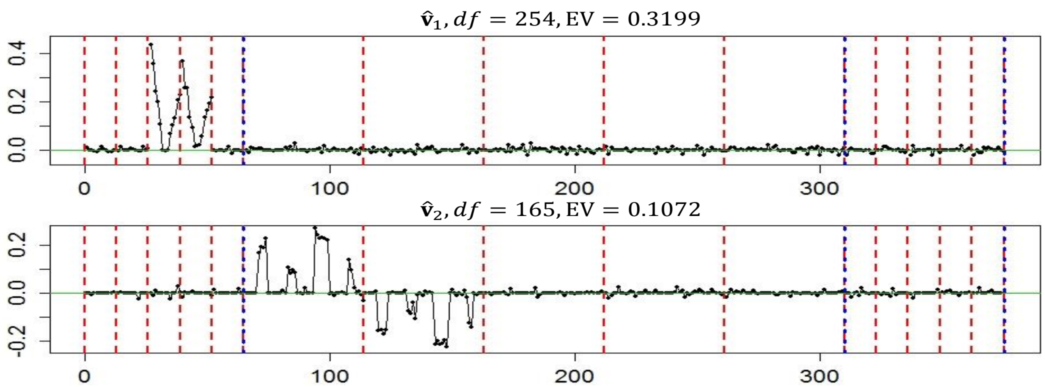

Figure S.7: Estimated eigenvectors by the ESMFPCA.

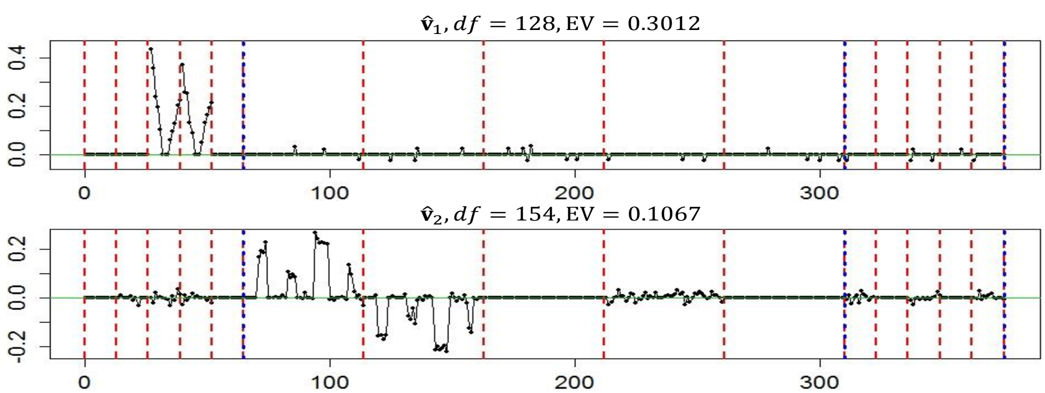

Figure S.8: Estimated eigenvectors by the PSMFPCA.

References

Akemann, G., Baik, J. and Di Francesco, P. (2011) *The Oxford Handbook of Random Matrix Theory*, Oxford University Press.

El Karoui, N. (2003) On the largest eigenvalue of Wishart matrices with identity covariance when n, p and p/n $\to \infty$. arXiv preprint math/0309355.

Ma, Z. (2013) Sparse principal component analysis and iterative thresholding. *The Annals of Statistics*, **41**(2), 772–801.