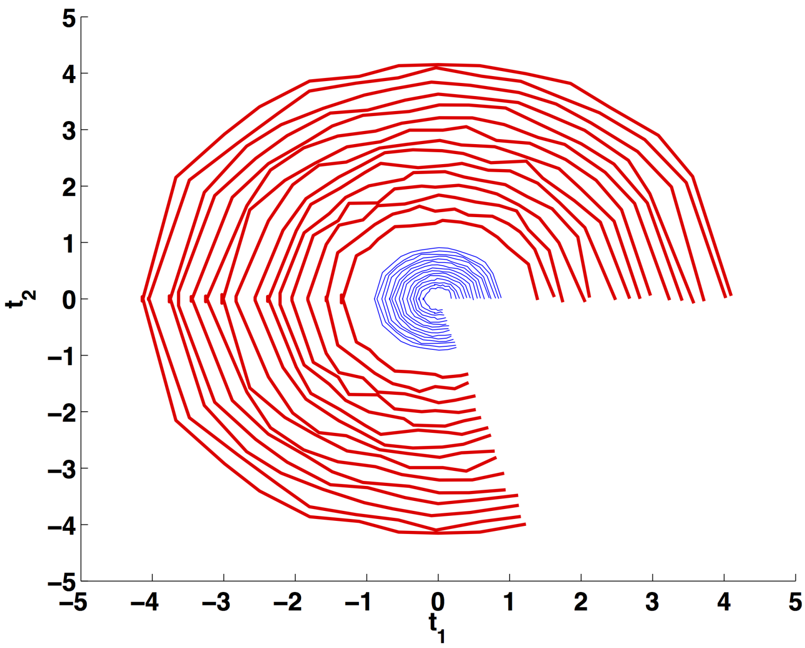
Pseudo-sample based contribution plots: innovative tools for fault diagnosis in kernel-based batch process monitoring

Supporting Material

Section 1 - Data simulation

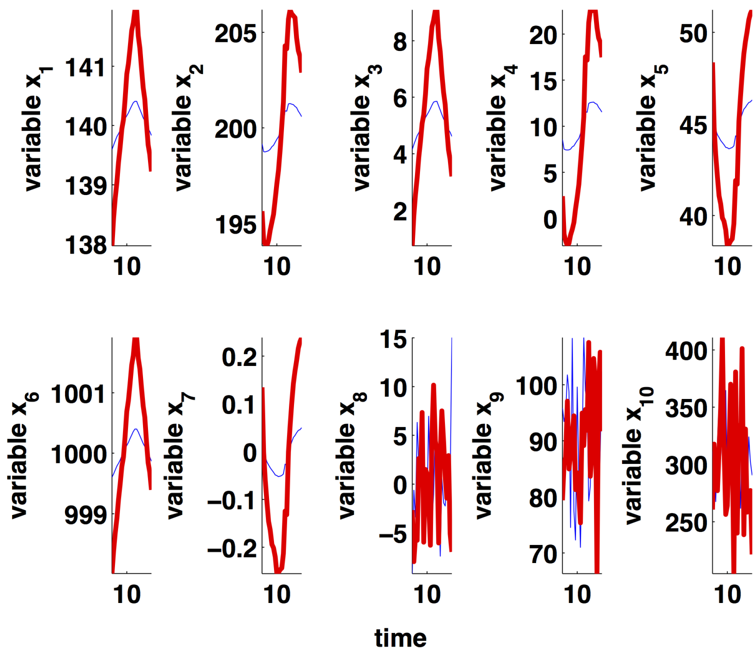
The data simulation was carried out generating a 750x2 set of scores, which define 2 classes of 15 different trimmed circular profiles of 25 observations each, as shown in SM.1.



SM. 1: On- (blue thin line) and off-specification (red thick line) simulated batch scores trajectories

Every trimmed circular profile represents a proper batch score trajectory, which may describe its progression in the latent variable space of a 2-

component PCA model built on a VWU data array. Multiplying this set of scores by a 2x7 transposed matrix of loadings (obtained performing PCA on real process data), a 750x7 dataset was constructed, which contains the trajectories of 7 variables in 30 different runs, constituted by 25 time samples each. Three noisy variables were also included for verifying the effectiveness of the pseudo-sample-based approach for diagnostic purposes.
As shown in SM. 2, the data simulation generated off-specification batches characterized by an increase in the variance of variables x_1-x_7, but not in their mean value with respect to the NOC ones.



SM. 2: Temporal evolution of the variables of the simulated dataset for a NOC (blue thin line) and a faulty (red thick line) batch

Section 2 - Variability increase detection case study

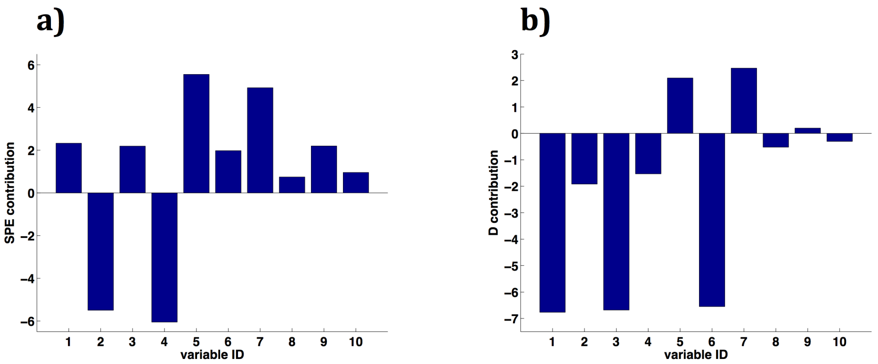
TCS-VWU

Table SM. 1: Overall Type I (OTI) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes. The table lists also the number of principal components (PCs) and the goodness of fit (R^2) of the two different models

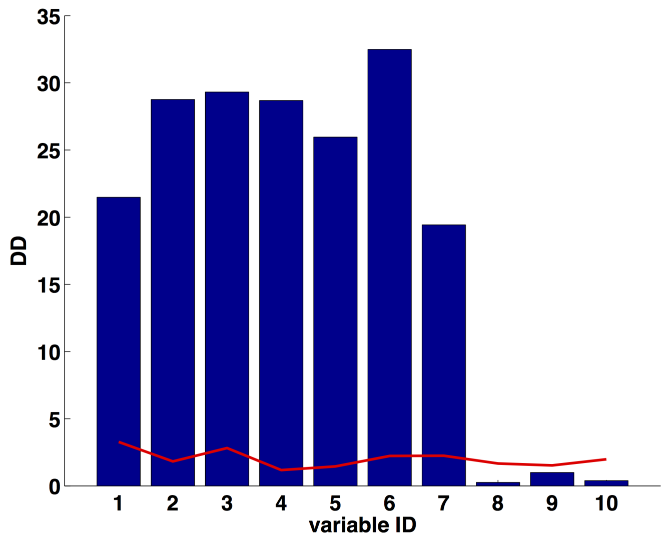
	PCs	R^2	$SPE_{ISL=5\%}$	$SPE_{ISL=1\%}$	$D_{ISL=5\%}$	$D_{ISL=1\%}$
Classical PCA	2	0.631	6.4%	1.6%	4.8%	1.6%
K-PCA (second-order polynomial)	2	0.729	6.4%	1.6%	5.6%	1.6%

Table SM. 2: Overall Type II ($OTII$) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes

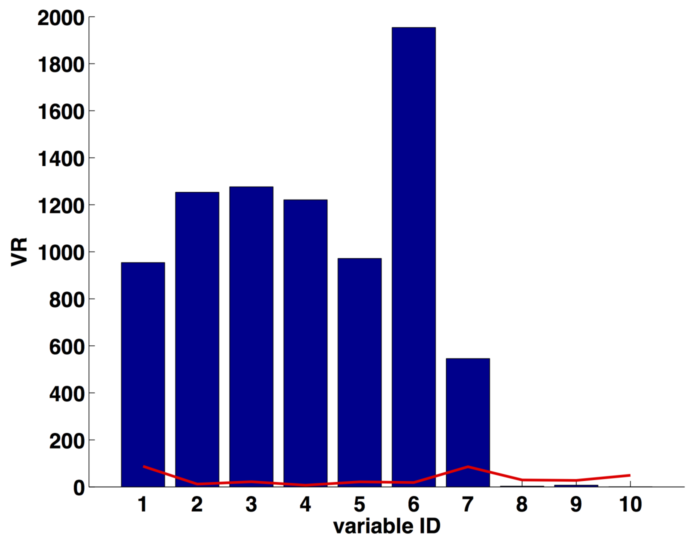
	$SPE_{ISL=5\%}$	$SPE_{ISL=1\%}$	$D_{ISL=5\%}$	$D_{ISL=1\%}$
Classical PCA	6.9%	13.6%	0.0%	0.0%
K-PCA (second-order polynomial)	0.0%	0.0%	0.0%	0.0%



SM. 3: Classical PCA: a) SPE and b) D-statistic contribution plots related to the first sampling point of the first faulty test batch. The displayed profiles are consistent with those observed for all the other *out-of-control* signals



SM. 4: K-PCA: Overall DD plot related to the first faulty test batch. The solid red line represents the 90% confidence limit calculated for the DD values of every variable by a jackknife-based procedure



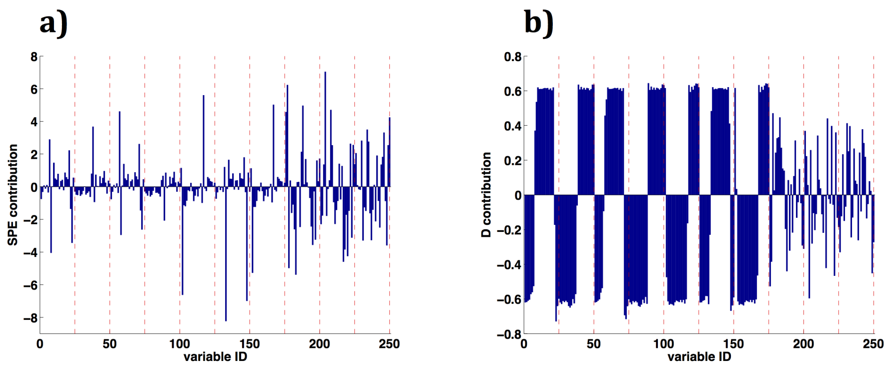
SM. 5: K-PCA: Overall VR plot related to the first faulty test batch. The solid red line represents the 90% confidence limit calculated for the VR values of every variable by a jackknife-based procedure

Table SM. 3: Type I (TI) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes. The table lists also the number of principal components (PCs) and the goodness of fit (R^2) of the two different models

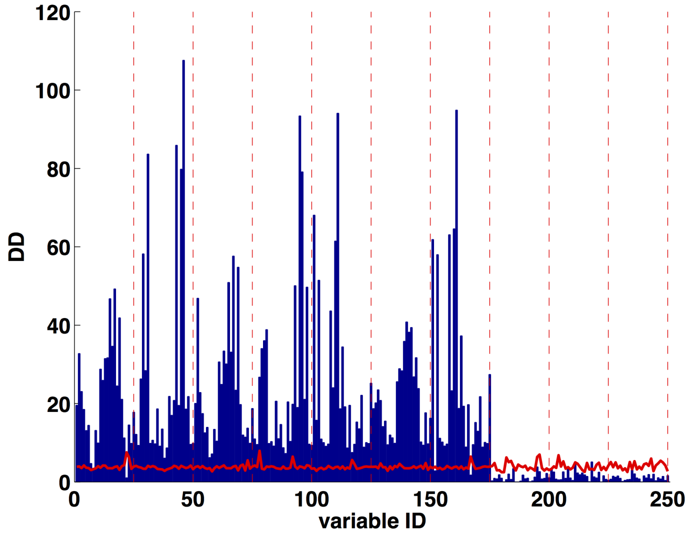
	PCs	R^2	$SPE_{ISL=20\%}$	$D_{ISL=20\%}$
Classical PCA	2	0.746	20.0%	20.0%
K-PCA (Gaussian, σ=89.1)	2	0.976	20.0%	20.0%

Table SM. 4: Type II (TII) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes

	$SPE_{ISL=20\%}$	$D_{ISL=20\%}$
Classical PCA	0.0%	0.0%
K-PCA (Gaussian, σ=89.1)	0.0%	0.0%



SM. 6: Classical PCA: a) SPE and b) D-statistic contribution plots related to the first faulty test batch. The vertical dotted red lines mark the separation between the contributions at the different sampling times of two consecutive process variables



SM. 7: K-PCA: Overall DD plot related to the first faulty test batch. The solid red line represents the 90% confidence limit calculated for the DD values of every variable at the different time instants by a jackknife-based procedure. The vertical dotted red lines mark the separation between the DD values at the different sampling times of two consecutive process variables

Section 3 - Variability decrease detection case study

VCS-VWU

Table SM. 5: Overall Type I (OTI) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes. The table lists also the number of principal components (PCs) and the goodness of fit (R^2) of the two different models

	PCs	R^2	$SPE_{ISL=5\%}$	$SPE_{ISL=1\%}$	$D_{ISL=5\%}$	$D_{ISL=1\%}$
Classical PCA	2	0.703	4.8%	1.6%	4.8%	1.6%
K-PCA (second-order polynomial)	2	0.697	4.8%	2.4%	6.4%	2.4%

Table SM. 6: Overall Type II ($OTII$) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes

	$SPE_{ISL=5\%}$	$SPE_{ISL=1\%}$	$D_{ISL=5\%}$	$D_{ISL=1\%}$
Classical PCA	96.5%	98.9%	51.7%	84.0%
K-PCA (second-order polynomial)	100%	100%	90.1%	100%

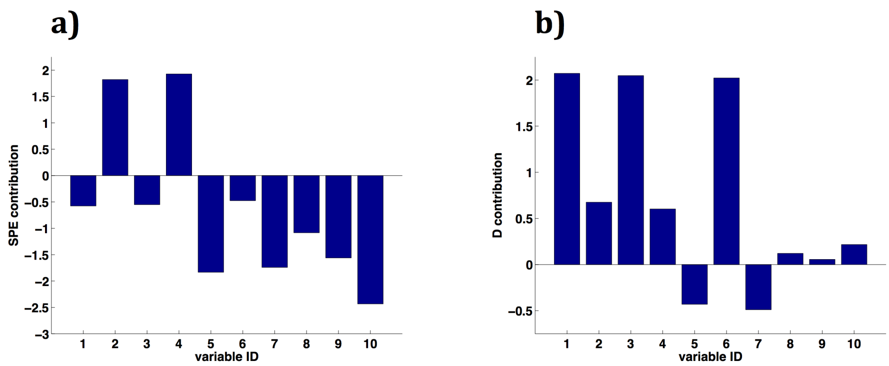
TCS-VWU

Table SM. 7: Overall Type I (OTI) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes. The table lists also the number of principal components (PCs) and the goodness of fit (R^2) of the two different models

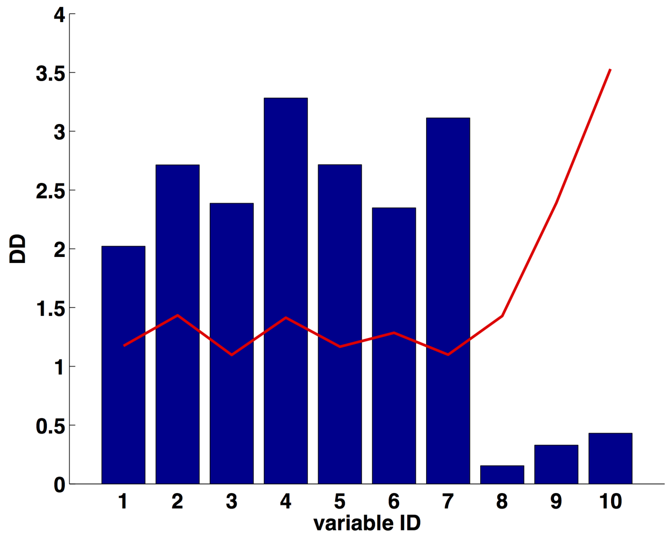
	PCs	R^2	$SPE_{ISL=5\%}$	$SPE_{ISL=1\%}$	$D_{ISL=5\%}$	$D_{ISL=1\%}$
Classical PCA	2	0.638	4.8%	1.6%	4.8%	3.2%
K-PCA (second-order polynomial)	3	0.812	5.6%	2.4%	4.8%	1.6%

Table SM. 8: Overall Type II ($OTII$) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes

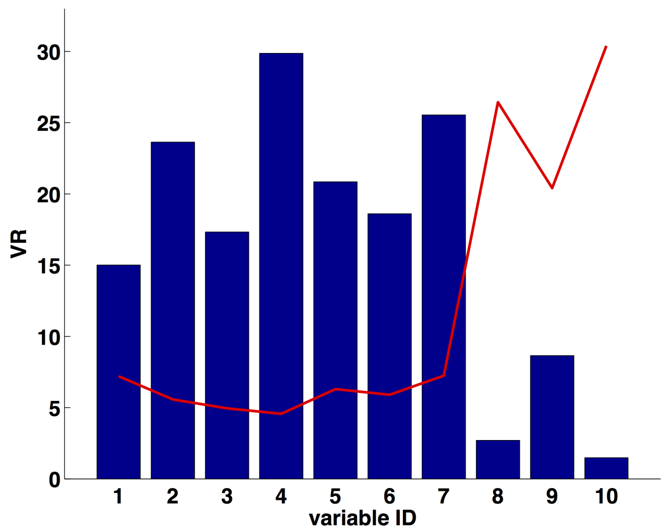
	$SPE_{ISL=5\%}$	$SPE_{ISL=1\%}$	$D_{ISL=5\%}$	$D_{ISL=1\%}$
Classical PCA	85.9%	91.7%	0.7%	24.5%
K-PCA (second-order polynomial)	56.3%	76.5%	0.0%	0.0%



SM. 8: Classical PCA: a) SPE and b) D-statistic contribution plots related to the first sampling point of the tenth faulty test batch. The displayed profiles are consistent with those observed for all the other *out-of-control* signals



SM. 9: K-PCA: Overall DD plot related to the tenth faulty test batch. The solid red line represents the 90% confidence limit calculated for the DD values of every variable by a jackknife-based procedure



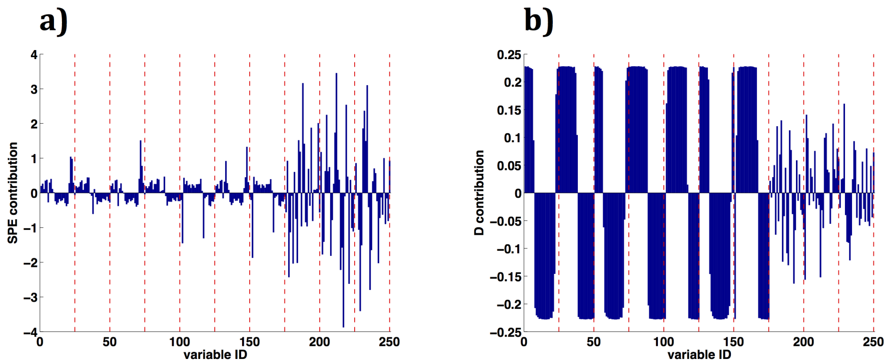
SM. 10: K-PCA: Overall VR plot related to the tenth faulty test batch. The solid red line represents the 90% confidence limit calculated for the VR values of every variable by a jackknife-based procedure

Table SM. 9: Type I (TI) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes. The table lists also the number of principal components (PCs) and the goodness of fit (R^2) of the two different models

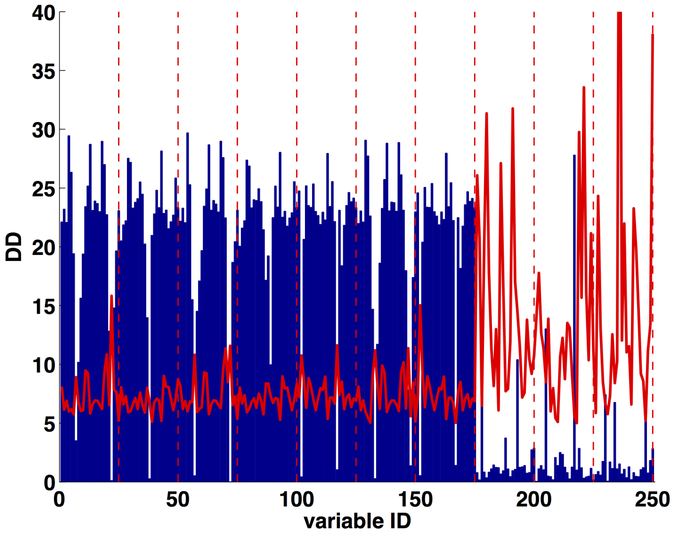
	PCs	R^2	$SPE_{ISL=20\%}$	$D_{ISL=20\%}$
Classical PCA	2	0.737	20.0%	20.0%
K-PCA (Gaussian, σ=19.3)	3	0.943	20.0%	20.0%

Table SM. 10: Type II (TII) risk values for the SPE and the D-statistic control charts resulting from both the classical PCA- and the K-PCA-based monitoring schemes

	$SPE_{ISL=20\%}$	$D_{ISL=20\%}$
Classical PCA	60.0%	0.0%
K-PCA (Gaussian, σ=19.3)	33.3%	0.0%

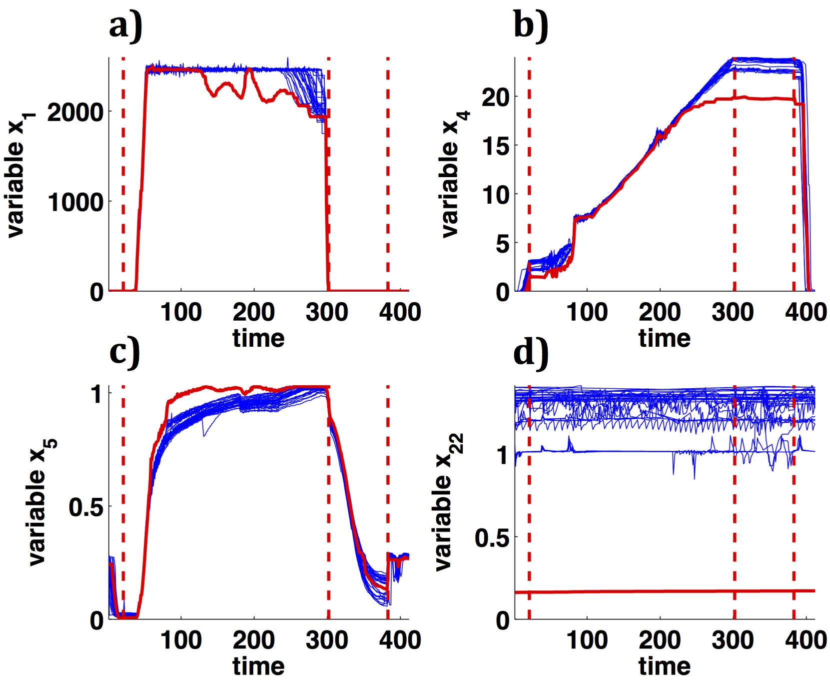


SM. 11: Classical PCA: a) SPE and b) D-statistic contribution plots related to the first faulty test batch. The vertical dotted red lines mark the separation between the contributions at the different sampling times of two consecutive process variables



SM. 12: K-PCA: Overall DD plot related to the first faulty test batch. The solid red line represents the 90% confidence limit calculated for the DD values of every variable at the different time instants by a jackknife-based procedure. The vertical dotted red lines mark the separation between the DD values at the different sampling times of two consecutive process variables

Section 4 - Further details on the chemical process dataset under study



SM. 13: Temporal evolution of the variables of the chemical process dataset, identified as having a high contribution to the fault, in the NOC batches (blue thin lines) and in the fifth faulty test process run (red thick line). The vertical dotted red lines mark the end point of every process stage. a) variable x_1; b) variable x_4; c) variable x_5; d) variable x_{22}