Supplementary Material for:

Beyond principal components: a critical comparison of factor analysis methods for subspace modelling in chemistry

Peter D. Wentzell, Cannon Giglio, and Mohsen Kompany-Zareh

Table S1. Comparison of figures of merit for different subspace estimation methods and noise types at a noise level of 0.1%.

Method	RMSE (measured)	RMSE (true)	Subspace Angle (U)	Subspace Angle (V)
Homoscedastic Errors (0.1%)				
	$\times 10^{-3}$	$\times 10^{-4}$	(◦)	(◦)
PCA	**0.9957±0.0057**	**2.503±0.061**	**0.3133±0.014**	**0.1356±0.015**
PCA (sc)	1.0159±0.0057	3.211±0.077	0.3617±0.016	0.1357±0.015
MLFA	0.9959±0.0057	2.512±0.061	0.3148±0.014	0.1356±0.015
PAF	0.9959±0.0057	2.513±0.062	0.3148±0.014	0.1356±0.015
MLPCA	**0.9957±0.0057**	**2.503±0.061**	**0.3133±0.014**	**0.1356±0.015**
MLPCA (est)	1.0055±0.0058	2.865±0.069	0.35912±0.017	0.1554±0.016
Column Homoscedastic Errors (0.1%)				
	$\times 10^{-4}$	$\times 10^{-4}$	(◦)	(◦)
PCA	**6.624±0.039**	1.807±0.037	0.2055±0.0086	0.08967±0.00933
PCA (sc)	6.673±0.039	1.643±0.037	0.1806±0.0072	0.08966±0.00932
MLFA	6.668±0.040	1.645±0.037	0.1807±0.0073	0.08966±0.00933
PAF	6.668±0.040	1.645±0.037	0.1807±0.0073	0.08966±0.00932
MLPCA	6.667±0.039	**1.640±0.036**	**0.1802±0.0073**	**0.08966±0.00932**
MLPCA (est)	6.731±0.039	1.880±0.042	0.2050±0.0081	0.10300±0.01047
Heteroscedastic Errors (0.1%)				
	$\times 10^{-4}$	$\times 10^{-5}$	(◦)	(◦)
PCA	**3.255±0.032**	9.316±0.292	0.10730±0.00640	0.04089±0.00494
PCA (sc)	3.282±0.032	8.821±0.302	0.09691±0.00575	0.04089±0.00493
MLFA	3.273±0.033	8.761±0.294	0.09662±0.00568	0.04089±0.00494
PAF	3.273±0.033	8.758±0.294	0.09659±0.00568	0.04089±0.00494
MLPCA	3.285±0.032	**8.216±0.300**	**0.09164±0.00554**	**0.03175±0.00354**
MLPCA (est)	3.317±0.032	9.392±0.342	0.10430±0.00629	0.03642±0.00419

Table S2. Comparison of figures of merit for different subspace estimation methods and noise types at a noise level of 5%.

Method	RMSE (measured)	RMSE (true)	Subspace Angle (U)	Subspace Angle (V)
Homoscedastic Errors (5%)				
	$\times 10^{-2}$	$\times 10^{-2}$	(◦)	(◦)
PCA	**4.981±0.024**	**1.257±0.028**	**15.54±0.60**	**7.037±0.672**
PCA (sc)	5.045±0.026	1.496±0.032	18.49±0.96	7.916±0.818
MLFA	4.982±0.024	1.261±0.028	15.61±0.60	7.049±0.672
PAF	4.983±0.024	1.263±0.028	15.62±0.61	7.178±0.701
MLPCA	**4.981±0.024**	**1.257±0.028**	**15.54±0.60**	**7.037±0.672**
MLPCA (est)	5.031±0.024	1.447±0.030	17.71±0.67	8.183±0.794
Column Homoscedastic Errors (5%)				
	$\times 10^{-2}$	$\times 10^{-3}$	(◦)	(◦)
PCA	**3.310±0.021**	9.142±0.198	10.283±0.432	4.718±0.473
PCA (sc)	3.335±0.021	8.281±0.207	8.994±0.413	**4.553±0.455**
MLFA	3.333±0.021	8.291±0.211	9.006±0.413	4.561±0.455
PAF	3.333±0.021	8.291±0.210	9.005±0.412	4.563±0.452
MLPCA	3.332±0.021	**8.269±0.206**	**8.973±0.413**	4.559±0.456
MLPCA (est)	3.364±0.021	9.470±0.246	10.192±0.458	5.291±0.592
Heteroscedastic Errors (5%)				
	$\times 10^{-2}$	$\times 10^{-3}$	(◦)	(◦)
PCA	**1.628±0.015**	4.698±0.015	5.330±0.337	2.109±0.246
PCA (sc)	1.642±0.015	4.438±0.014	4.804±0.285	2.084±0.244
MLFA	1.638±0.015	4.404±0.014	4.793±0.293	2.086±0.247
PAF	1.638±0.015	4.402±0.014	4.792±0.293	2.086±0.246
MLPCA	1.644±0.015	**4.127±0.014**	**4.538±0.301**	**1.590±0.192**
MLPCA (est)	1.660±0.015	4.709±0.018	5.150±0.366	1.797±0.214

Table S3. Comparison of figures of merit for different subspace estimation methods and noise types at a noise level of 10%.

Method	RMSE (measured)	RMSE (true)	Subspace Angle (U)	Subspace Angle (V)
Homoscedastic Errors (10%)				
	$\times 10^{-1}$	$\times 10^{-2}$	(◦)	(◦)
PCA	**0.9949±0.0057**	**2.570±0.061**	**29.48±1.22**	**15.42±1.52**
PCA (sc)	1.0043±0.0063	2.950±0.139	38.11±6.18	22.04±6.75
MLFA	0.9951±0.0057	2.579±0.061	29.66±1.24	15.53±1.49
PAF	0.9974±0.0096	2.660±0.245	34.03±14.95	21.52±18.04
MLPCA	**0.9949±0.0057**	**2.570±0.061**	**29.48±1.22**	**15.42±1.52**
MLPCA (est)	1.0051±0.0057	2.967±0.073	33.33±1.53	18.40±1.81
Column Homoscedastic Errors (10%)				
	$\times 10^{-2}$	$\times 10^{-2}$	(◦)	(◦)
PCA	**6.608±0.043**	1.856±0.040	20.86±0.88	10.80±1.31
PCA (sc)	6.659±0.043	1.655±0.037	17.67±0.75	**9.35±1.10**
MLFA	6.655±0.044	1.660±0.037	17.69±0.76	9.41±1.12
PAF	6.655±0.044	1.659±0.038	17.69±0.76	9.40±1.12
MLPCA	6.653±0.043	**1.655±0.037**	**17.64±0.74**	9.40±1.12
MLPCA (est)	6.717±0.043	1.905±0.045	20.02±0.93	10.99±1.25
Heteroscedastic Errors (10%)				
	$\times 10^{-2}$	$\times 10^{-3}$	(◦)	(◦)
PCA	**3.249±0.042**	9.405±0.308	10.775±0.617	4.459±0.597
PCA (sc)	3.276±0.042	8.842±0.314	9.637±0.552	4.245±0.544
MLFA	3.269±0.043	8.786±0.318	9.627±0.557	4.264±0.550
PAF	3.269±0.043	8.782±0.318	9.623±0.556	4.258±0.550
MLPCA	3.280±0.042	**8.215±0.300**	**9.071±0.525**	**3.161±0.445**
MLPCA (est)	3.312±0.043	9.410±0.339	10.332±0.674	3.698±0.513

Supplementary Figure S1:

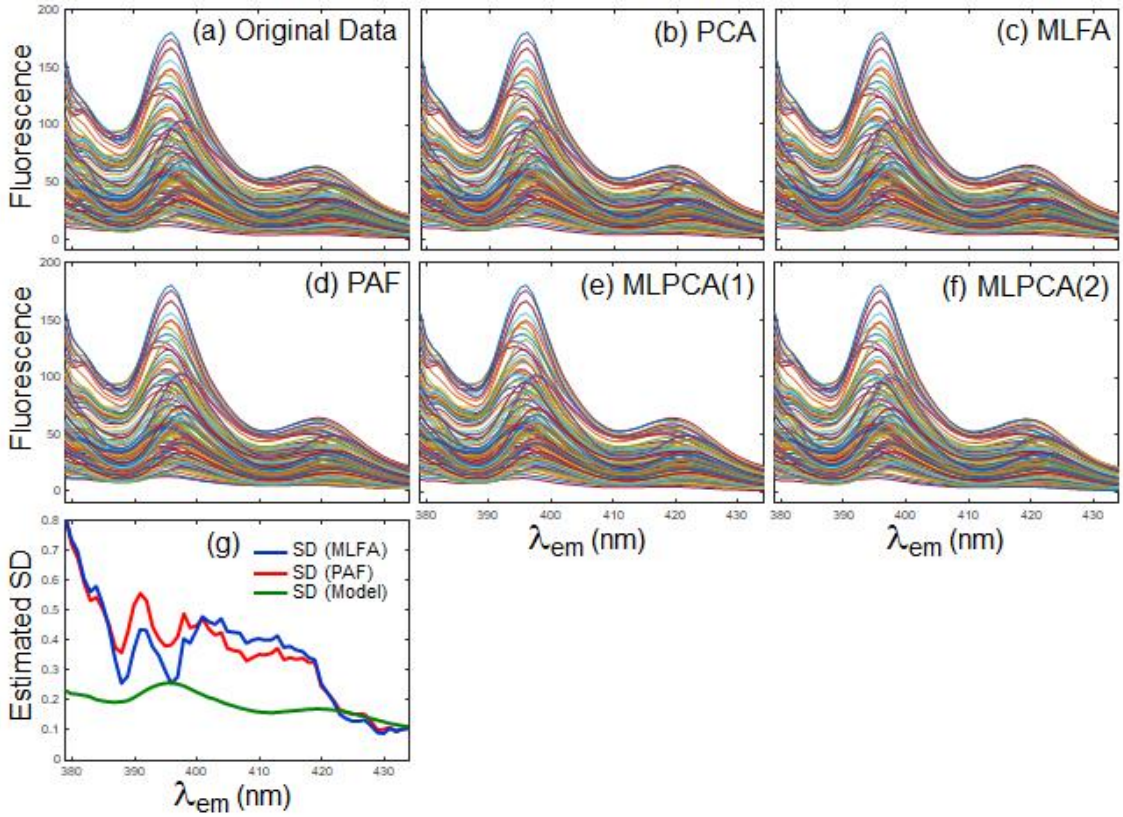

Figure S1. Comparison of original and reconstructed PAH fluorescence emission spectra for region excluding Rayleigh scatter (λ_{ex} = 300-374 nm, λ_{em} = 379-434 nm). (a) Original fluorescence emission spectra, with spectra reconstructed using (b) PCA, (c) MLFA, (d) PAF, (e) MLPCA with homoscedastic errors, and (f) MLPCA using measurement error model. Panel (g) shows the measurement uncertainty estimates from MLFA (blue) and PAF (blue) as well as the pooled error model used for MLPCA.

Supplementary Figure S2:

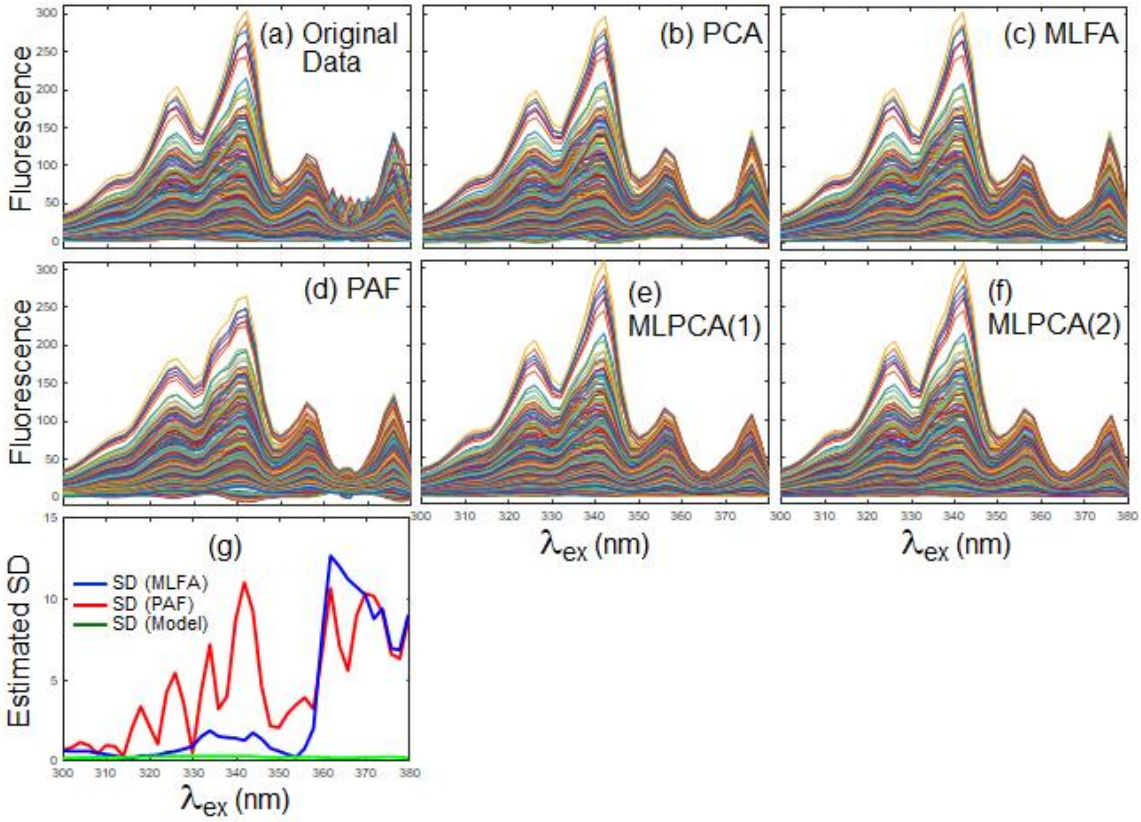

Figure S2. Comparison of (a) original fluorescence excitation spectra with spectra reconstructed using three-component models for (b) PCA, (c) MLFA, (d) PAF, (e) MLPCA using uniform measurement errors, and (f) MLPCA using measurement error model. Panel (g) shows the measurement uncertainty estimates from MLFA (blue) and PAF (blue) as well as the pooled error model used for MLPCA.

Table S4. Root-mean-squared (RMS) differences among fluorescence excitation spectral data reconstructed using various methods.

Method	PCA	PCA(sc)	MLFA	PAF	MLPCA(1)
PCA (sc)	4.38				
MLFA	2.38	5.09			
PAF	3.71	2.58	4.15		
MLPCA(1)	3.36	5.26	2.43	4.54	
MLPCA(2)	3.44	5.21	2.56	4.48	0.68