Supporting Information

Operando UV-Vis Spectroscopy for Real-Time Monitoring of Nanoparticle Size in Reaction Conditions: A Case Study on r-WGS over Au Nanoparticles

Chiara Negri,[a] Riccardo Colombo,[a] Mauro Bracconi,[a] Cesare Atzori,[b] Alessandro Donazzi,[a] Andrea Lucotti,[c] Matteo Tommasini[c] and Matteo Maestri*[a]

1. Additional experimental results

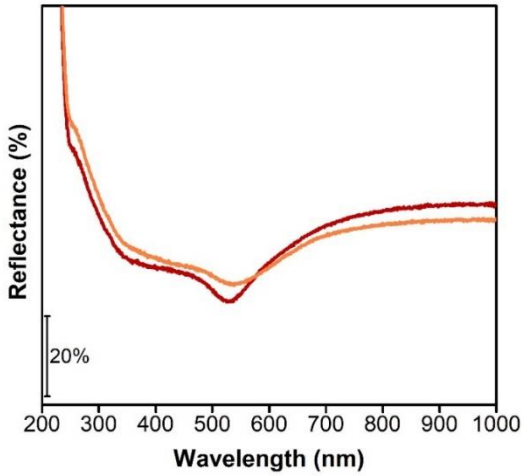

Figure S1. ex-situ UV-Vis DR spectra of $AuAl_2O_3$_30 (red curve) and $AuAl_2O_3$_60 (orange)

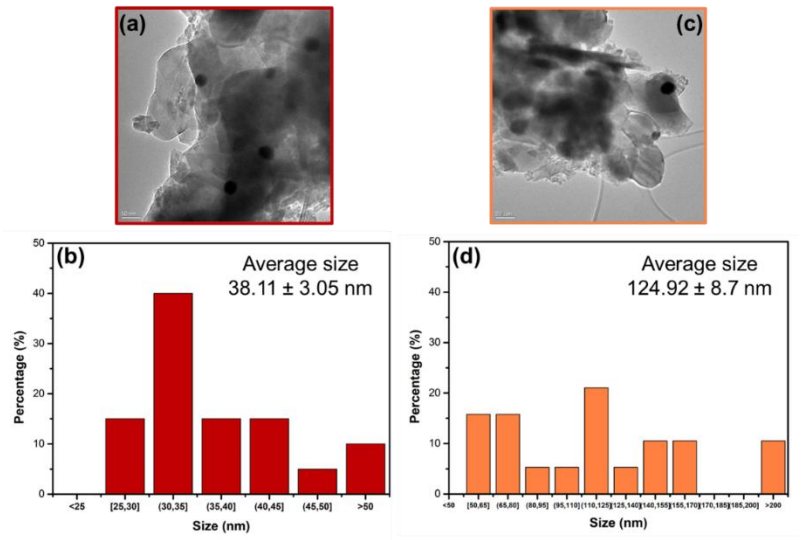

Figure S2. (a) HR-TEM image and (b) NPs size distribution of $AuAl_2O_3$_30; (c) HR-TEM image and (d) NPs size distribution of $AuAl_2O_3$_60

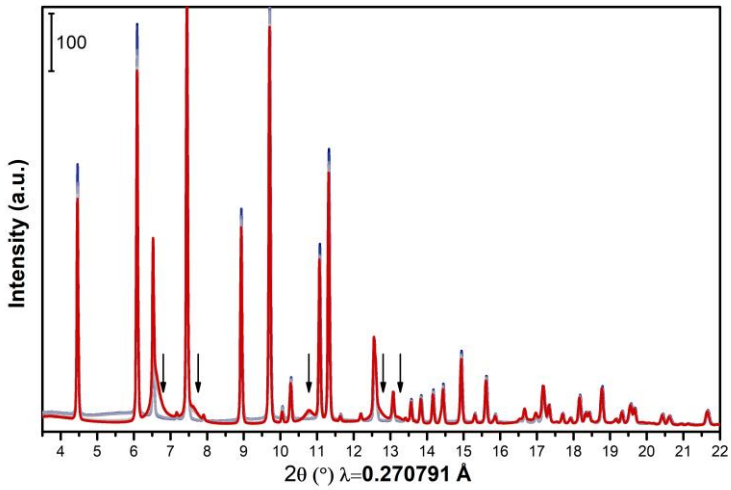

Figure S3. *operando* PXRD patterns at λ=0.270791 Å during H2-pretreatment from RT to 200°C (from blue to red curves), with highlighted Au NPs reflexes

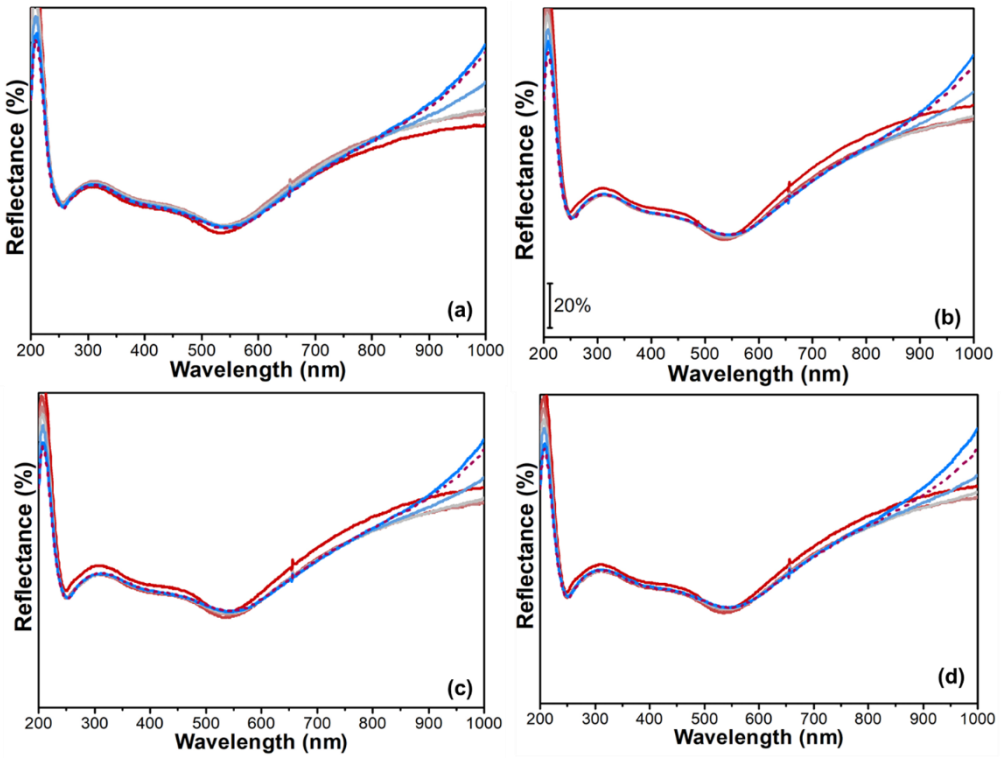

Figure S4. Operando UV-Vis DR spectra of the Au/Al_2O_3_4 sample during rWGS from RT to 600°C (from red to light blue curves) and after 230 minutes in isotherm at 600°C (dashed purple curve): (a) First test, (b) second test, (c) third test and (d) fourth test. The spectra were acquired while feeding 5%CO_2/5% H_2 – N_2 balanced.

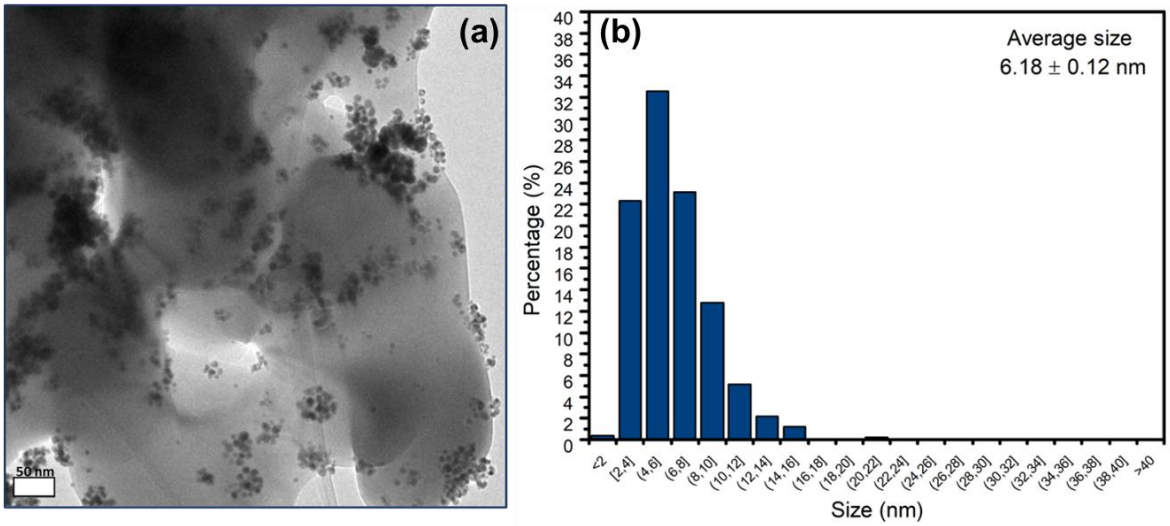

Figure S5. (a) HR-TEM image of the spent $AuAl_2O_3$_4 catalyst and (b) NPs average size distribution

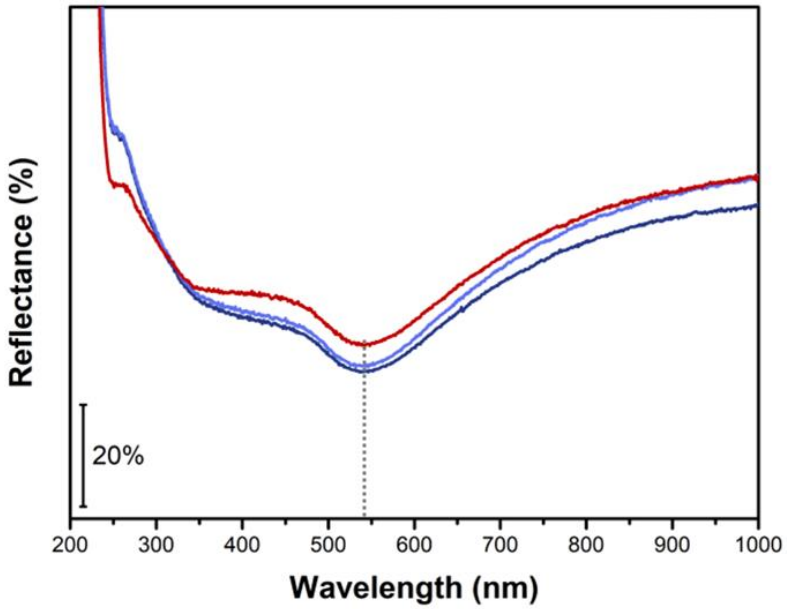

Figure S6. Ex situ UV-Vis DR spectra of the fresh Au/Al_2O_3_4 catalyst (dark red curve), spent Au/Al_2O_3_4 after the first rWGS run (light blue curve) and spent Au/Al_2O_3_4 after the end of the last rWGS run (dark blue curve).

2. Determination of the number of pure components via Principial Component Analysis (PCA)

The first step towards the MCR-ALS analysis is the determination of the number of pure components (N_{PC}) to explain the variance observed in the experimental dataset above the noise level. The experimental dataset D consists of mxn matrix where m is the number of energy points and n is the number of spectra.
To this aim, the Principal Component Analysis (PCA) of the XANES of Au (H_2 reduction from room temperature to 200 °C) has been performed.
The PCA is computed by using the Singular Value Decomposition (SVD) of the experimental XANES dataset D which decomposes the data matrix in the product of three matrices.

$$D = U\Sigma V^T \quad \text{(S1)}$$

where U and V are unitary square matrices while Σ is a diagonal matrix. Matrix U contains along the columns the eigenvectors of the covariance matrix associated with D, i.e. $Z = DD^T$, whereas matrix V^T contains along

the rows the eigenvectors of the transpose of Z. The diagonal elements of S are named singular values and they are connected to the eigenvalues of Z through Eq.(S2).

$$\lambda_i = \frac{s_i^2}{m-1} \tag{S2}$$

where s_i and λ_i and the singular values and the eigenvalues corresponding to S and to the data covariance matrix Z, respectively.

The best number of principal components (PC), that corresponds to the number of pure components, has been obtained by combining qualitative analysis along with statistical tests, i.e. Imbedded Error (IE) function, IND-factor and Malinowski F-Test[1].

The qualitative analysis is rooted on the analysis of the abstract components and of the R-factor. The abstract components that represent the columns of the matrix obtained by multiplying the principal components to the corresponding singular values.

$$R = U\Sigma \tag{S3}$$

The abstract components do not have any chemical meaning and correspond a mathematical solution of Eq. (S**1**) but they help in discriminating between relevant features and noise-related components. A second analysis is based on the analysis of the reconstruction residuals and the R-factor (Eq. (S4))

$$R_{factor,i}^{k} = \frac{\left\| D - R_k V_k^T \right\|}{\|D\|} \tag{S4}$$

where k corresponds to the number of principal components considered for the reconstruction. When the dataset is properly reproduced, the R-factor evaluated on each scan must have comparable magnitude.

The scree plot reports the variance or the singular value magnitude associated to each principal components against their number. The variance associated to the components rapidly decreases up to an elbow that indicate the border between components retaining a chemical meaning from those related to the noise, whose contribution to the dataset is similar resulting to lie on a quasi-flat line.

The IE function is given by the following expression:

$$IE = \sqrt{\frac{k\sum_{i=k+1}^{n}\lambda_i}{mn(n-k)}} \tag{S5}$$

where k corresponds to the number of components used to reproduce D. Under the assumption that the experimental errors are randomly and uniformly distributed, the IE function progressively decreases until it reaches a minimum and then slowly increases. The position of the minimum corresponds to the number of correct PCs.

The IND-factor has been proposed by Malinowski to provide a more sensitive function to properly select the number of PCs. It is based on the following empirical function:

$$IND = \frac{1}{(n-k)^2}\sqrt{\frac{\sum_{i=k+1}^{n}\lambda_i}{m(n-k)}} \tag{S6}$$

The function shows a behavior with a minimum as the IE function, but it has been observed that the minimum is usually more evident and it also appears when the IE function fails.

The Malinowski F-Test it is based on the evidence that the reduced eigenvalues $\tilde{\lambda}_i$ are constant for non-significant PCs:

$$\tilde{\lambda}_i = \frac{\lambda_i}{(m-i+1)(n-i+1)} \tag{S7}$$

Since the $\tilde{\lambda}_i$ are still proportional to a variance, it is possible to employ a Fischer test. The test starts from the smallest $\tilde{\lambda}_i$, associated with the noise, and proceeds towards the $\tilde{\lambda}_i$ with higher magnitude until the first

significant one is obtained. The kth component is considered significant on the bases of a F-Test applied on the related standardized F-variable:

$$F(1, n-k) = \frac{\tilde{\lambda}_i}{\sum_{i=k+1}^{n} \lambda_i} \left(\sum_{i=k-1}^{n} (m-i+1)(n-i+1) \right) \quad \text{(S8)}$$

When the percentage of significance level associated, with the kth F-variable, is lower than a pre-defined threshold value, i.e. 5%, then the kth is accepted as a pure component.

We have examined the abstract components for the Au sample to assess the significance of the components. Figure S7 reports the first four abstract components for the Au XANES dataset.

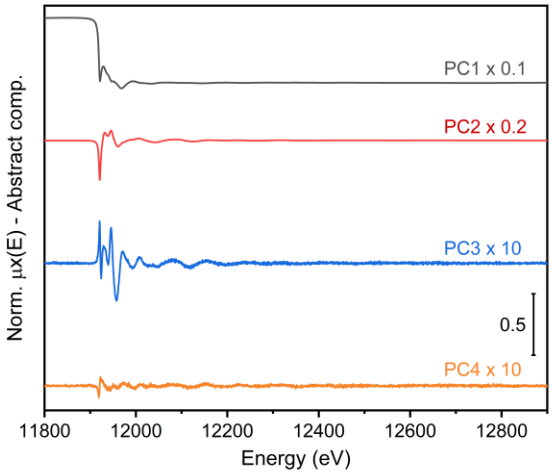

Figure S7. First four abstract components determined by PCA analysis on the Au XANES dataset

The first abstract component retains most of the variance and it has a shape similar to an inverted XANES spectrum as already reported in literature[1]. The second component captures the second highest spectral features whereas the third describes minor variations of the spectrum. Starting with the fourth components, they appear featureless for the description of the spectra and they follow a trend as a function of the energy that is typical of instrumental noise.

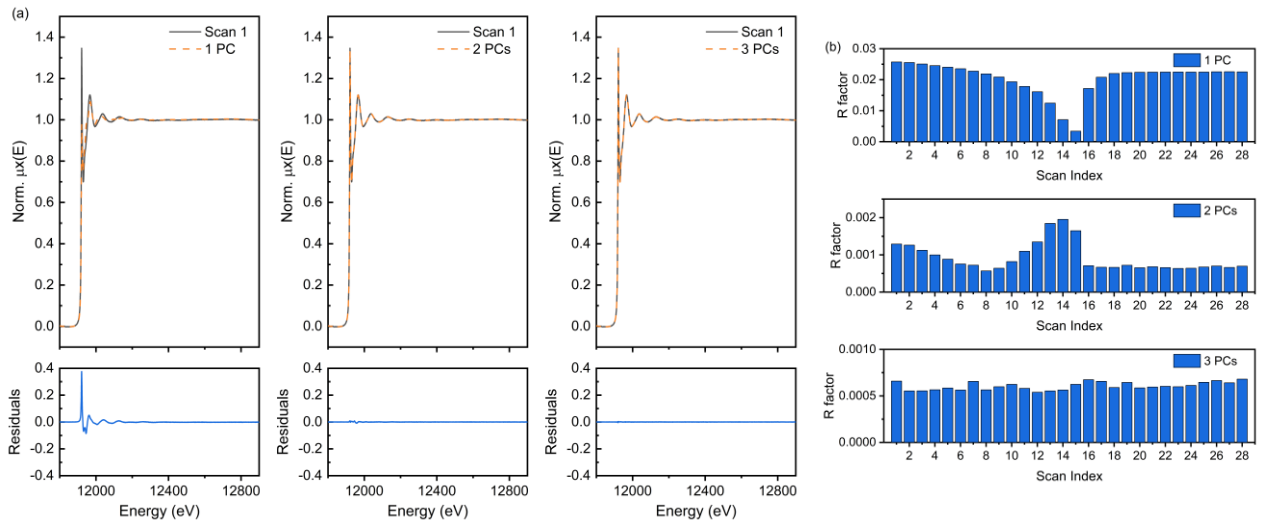

Figure S8. (a) Reconstruction of the first scan using one to three components along with the residuals as a function of energy; (b) R-factor for each scan as a function of the number of PCs

A further validation of the results can be obtained by observing the residuals and the R-factor related to the reconstruction. It can be noted that by increasing the number of PCs the quality of the reconstruction

increases as expected. With one PC, there is a lack of reproduction connected with the white line feature as shown by the residuals plot (see Figure S8(a)). When two PCs are employed, the reconstruction strongly improves but the R-factor trend is still not homogenous, as shown in Figure S8(b), whereas the reproduction with 3 PCs becomes excellent.

Albeit informative, the previous analyses are qualitative. Hence, we evaluated the number of PCs by considering additional analysis based on statistical tests.

The scree plot for the Au XANES dataset reveals the elbow is located at the third component as shown in

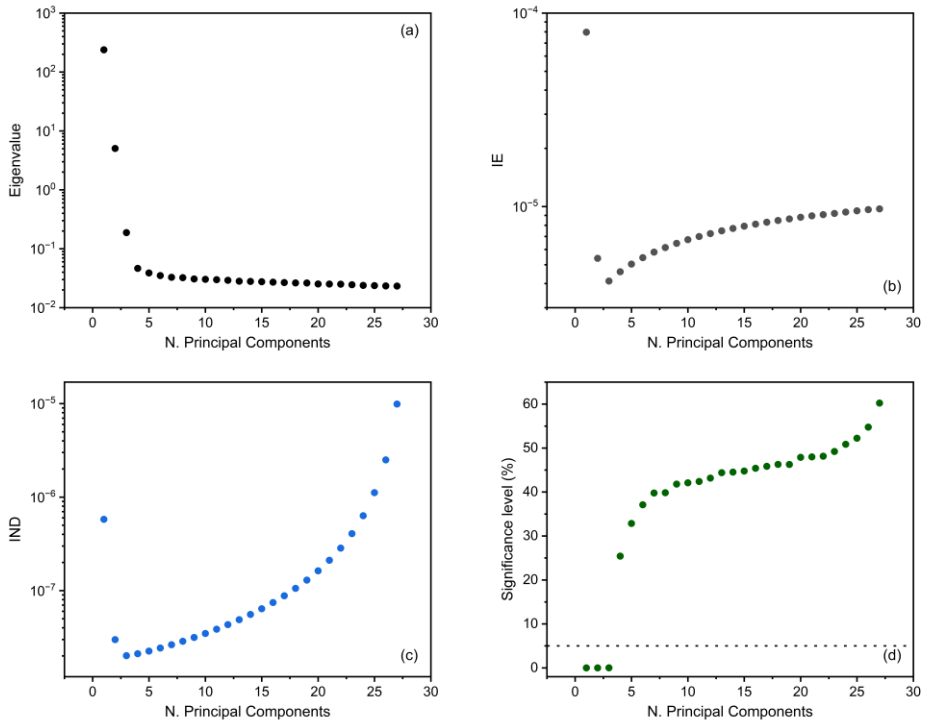

Figure S9. Statistical estimators to evaluate the number of principal components contributing to the Au XAS dataset: (a) scree plot; (b) Imbedded Error (IE) function; (c) IND factor; (d) Malinowski F-test

Figure S9(a). A further increase in the number of components provides a modest variation of the eigenvalues and after the 6th component we observe a full stabilization of the eigenvalues. The IE function and the IND-factor reveal an evident minimum in correspondence of three PCs confirming the qualitative information from the scree plot as shown in Figure S9(b)-(c). Moreover, the Malinowki F-Test highlights that three PCs are sufficient to describe the system as well as reported in Figure S9(d).

Based on the phenomenological evidence and on the quantitative information from the tests, we have selected $N_{PC} = 3$ for the subsequent analysis.

3. MCR-ALS details on the methodology and implementation

Multivariate Curve Resolution is a potent chemometric algorithm that enables to decompose an experimental dataset of spectra D, consisting of n spectra each measured at m energy points, into pure contributions consisting of a set of spectra S of different chemical compounds, i.e. pure components, and their corresponding concentrations C. Mathematically, this corresponds to the following bilinear decomposition:

$$D = SC^T + E \tag{S9}$$

where E is the matrix of the residuals not explained by the model.

Several MCR algorithms have been proposed in the literature based on non-iterative and iterative approaches[2,3]. Among those, the Multivariate Curve Resolution – Alternate Least Squares (MCR-ALS) is an

iterative method proposed by Tauler and co-worker[4,5] and widely employed in the last two decades for spectral unmixing.

The method requires to define the number of pure components by exploiting a physical-chemical understanding of the system or by dimensionality reduction techniques such as the PCA (see Section 3 for further details).

Then, it is necessary to have an initial estimate of either the concentration profiles or the spectra of pure components. There are many possibilities to generate initial estimates, but random estimates should be avoided to facilitate convergence. Efficient methods to generate the initial estimates are the evolving factor analysis (EFA)[6] and the simple-to-use interactive self-modeling mixture analysis (SIMPLISMA) algorithm[7]. The EFA algorithm monitors the gradual change in the number of significant components during the measurement by performing local rank analysis of sub-dataset of increasing size. As outcome of the analysis, the concentration profiles are obtained and they can be employed as initial guess for the ALS algorithm. On the other hand, the SIMPLISMA algorithm works selecting in a sequential way the variables that have less information in common with the previously selected ones.

attempts to identify a set of pure spectral profiles equal to the number of significant PCs.

Once the initial estimate is available, the method MCR-ALS method consists of iteratively solving the following two least-squares problems until convergence:

$$\min_{C} \|D - SC^T\| \tag{S10}$$

$$\min_{S} \|D - SC^T\| \tag{S11}$$

The main weakness of the MCR-ALS algorithm stands in the rotational and intensity ambiguity resulting in non-unique solution of the decomposition problem. A common approach to mitigate or even suppress this problem is the usage of constraints in the algorithm that can also guarantee physical-chemical meaningfulness of the results. The most common soft constraints that can be either applied to S and C matrices are: (i) *non-negativity*, force the profiles to be positive; (ii) *unimodality*, force the profile to have a single maximum; (iii) *monotonicity*, force the profiles to be increasing (or decreasing); (iv) *closure*, force mass conservation in the system. Hard constraints are not currently implemented in the framework. Hence, in each iteration, S and C matrices obtained from Eqs. (S10)-(S11) are corrected to obey to the constraints obtaining the $\hat{S}$ and $\hat{C}$ matrices.

The $\hat{S}$ and $\hat{C}$ matrices enable to reconstruct the initial dataset and to evaluate the residual matrix E:

$$E = D - \hat{S}\hat{C}^T \tag{S12}$$

The residual matrix is employed to compute the lack-of-fit LOF:

$$LOF = \frac{\|E\|}{\|D\|} \tag{S13}$$

The procedure iterates unless the LOF between consecutive iterations does not significantly improve (i.e. it is below a user-defined tolerance usually equal to 0.1%). However, it is also possible that the LOF starts to increase, indicating possible non-convergence of the procedure. In this case, the algorithm is automatically stopped and the results have to be carefully analyzed. At the end of the procedure, the explained variance Eq. (S14) and the standard deviation of residuals with respect to experimental data are also evaluated to quantify the quality of the unmixing.

$$R^2 = \frac{\|D\|^2 - \|E\|^2}{\|D\|^2} \tag{S14}$$

$$\sigma = \frac{\|E\|^2}{mn} \tag{S15}$$

where d_{ij} and $e_{ij}$are the elements of D and E, respectively.

4. MCR-ALS results for the Au XANES dataset

We have employed the MCR-ALS procedure to unmix the spectral dataset regarding the Au XANES dataset by considering $N_{PC} = 3$, i.e. the correct value provided by the statistical analysis. The ALS routing is initialized by using the SIMPLISMA algorithm. Soft constraints have been employed: (i) non-negativity for both spectra and concentrations and (ii) closure to 1 for the concentration to enforce mass conservation.

Table S1. Quality indicators of MCR-ALS analysis ($N_{PC} = 3$) of the Au XANES dataset

MCR-ALS Quality Indicator	Value
Std. deviation of residual vs exp. data	$6.5 \cdot 10^{-4}$
Fitting error (LOF) of D_{PCA} (%)	$3.3 \cdot 10^{-2}$
Fitting error (LOF) of D (%)	$6.9 \cdot 10^{-2}$

The MCR-ALS algorithm successfully converged (i.e. LOF variation between two consecutive iterations <0.1%) in 23 iterations. Table S1 lists a series of indicators for the achieved reconstruction of the dataset proving the overall quality of the decomposition.

5. MCR-ALS results for the Au XANES dataset for different values of N_{PC}

The MCR-ALS procedure has been also employed in a downsized PC space around the optimal dimension identified by statistical analysis. From the qualitative analysis of Section 2, the minimum number of PCs that is able to account for the key modifications in the spectral dataset is obtained for $N_{PC} = 2$. The MCR-ALS algorithm has been employed by keeping constant the same parameters and constraints used for the optimal $N_{PC} = 3$. The MCR-ALS converged in 6 iterations with a fitting error (LOF) of D of $6.9 \cdot 10^{-1}$ % and a standard deviation of the residuals with respect to experimental data of $9.8 \cdot 10^{-4}$.

Figure S10 compares the *pure* components spectra obtained with the downsized PC spaces with the results obtained with optimal number of PCs, discussed in the main text (Section 3.2). It is evident that, in the case of $N_{PC} = 2$, the minimal set of Au-species includes (i) an initial oxidized Au^{3+} species and (ii) a reduced Au^0 obtained at the end of the reduction procedure as already observed for the optimal number of PCs, whereas the intermediate species is lost.

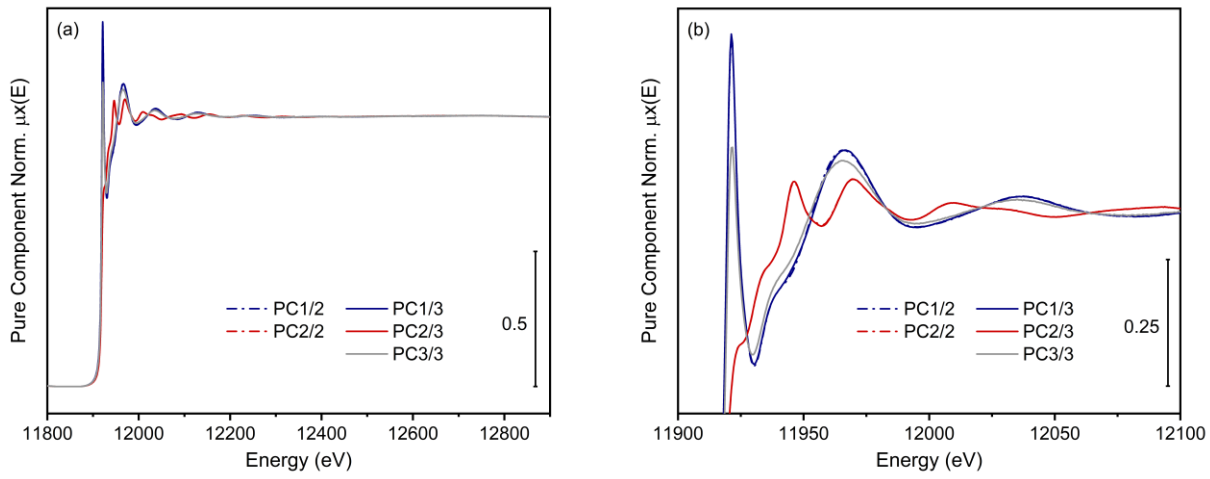

Figure S10. (a) XANES spectra of pure components derived from MCR-ALS for $N_{PC} = 2$ (dash dot lines) and $N_{PC} = 3$ (solid lines); (b) zoom of the white line region

The variation of the N_{PC} does not result in a strong modification of the pure spectra since by incrementing the number of PCs from 2 to 3 the curves are unchanged as shown in Figure S10(b). That confirms the stability of the reconstruction algorithm as well as the robustness of the first guess evaluated using the SIMPLISMA method.

Figure S11 reports the concentration profiles of pure components for the optimal and downsized number of PCs. It is possible to observe a significant rearrangement of the concentration profiles reducing the number of PCs. The effect of the elimination of PC3 determines a slower consumption of PC1, i.e. Au^{3+}, and a concomitant slower formation of PC2, i.e. Au^{0}. This trend seems to reproduce with less accuracy experimental evidences of both XANES and UV-Vis operando spectra.

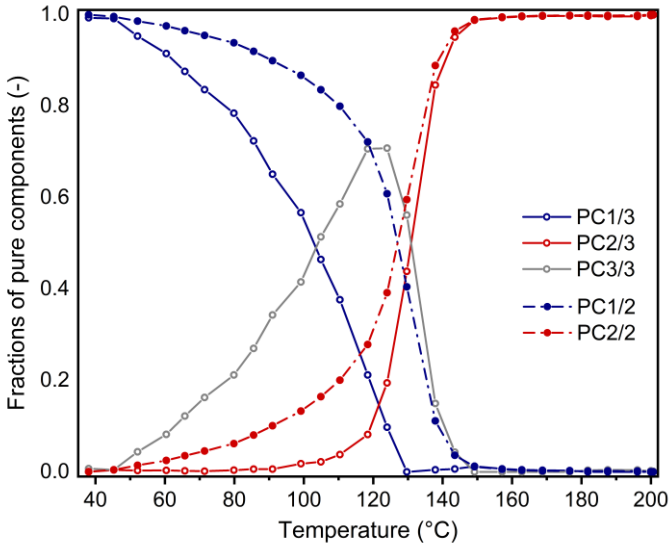

Figure S11. Temperature dependent concentration profiles of pure component derived from MCR-ALS for $N_{PC} = 2$ (dash dot lines and full circle) and $N_{PC} = 3$ (solid lines and empty circle)

6. References

1 A. Martini and E. Borfecchia, *Crystals*, 2020, 10, 1–46.
2 A. De Juan, S. C. Rutan and R. Tauler, *Compr. Chemom.*, 2009, **2**, 325–344.
3 Y. Z. Liang, *Compr. Chemom.*, 2009, **2**, 309–323.
4 R. Tauler, *Chemom. Intell. Lab. Syst.*, 1995, **30**, 133–146.
5 A. De Juan and R. Tauler, *Anal. Chim. Acta*, 2003, **500**, 195–210.
6 M. Maeder, *Anal. Chem.*, 1987, **59**, 527–530.
7 W. Windig, C. E. Heckler, F. A. Agblevor and R. J. Evans, *Chemom. Intell. Lab. Syst.*, 1992, **14**, 195–207.