Supporting Information

The Correlation between Multiple Variable Factors and Autocatalytic Property of Cerium Oxide Nanoparticles based on Cell Viability

Changyan Li [a,b], Xiangcheng Shi [a,b], Liang Bao [c,d], Jingjie Yang [a,b], Alatangaole Damirin [c], Jun Zhang [a,b*]

[a] College of Chemistry and Chemical Engineering, Inner Mongolia University, Hohhot 010021, P. R. China

[b] Inner Mongolia Key Lab of Nanoscience and Nanotechnology, Inner Mongolia University, Hohhot 010021, P. R. China

[c] College of Life Sciences, Inner Mongolia University, Hohhot 010021, P. R. China

[d] College of Biochemical Engineering, Hohhot Vocational College, Hohhot, 010051, P. R. China

* Corresponding author:

Jun Zhang, Ph.D.

College of Chemistry and Chemical Engineering,

Inner Mongolia University, Hohhot 010021, P. R. China

Tel: 86-471-4995400

E-mail: cejzhang@imu.edu.cn

Figure Captions

Figure 1. The functional relationship between particle size and lattice parameter of Ce NPs (111) lattice plane for S1, S2 and S3 samples

Figure 2. The functional relationship between particle size and atomic ratio of Ce^{3+} in S1, S2 and S3 Ce NPs samples

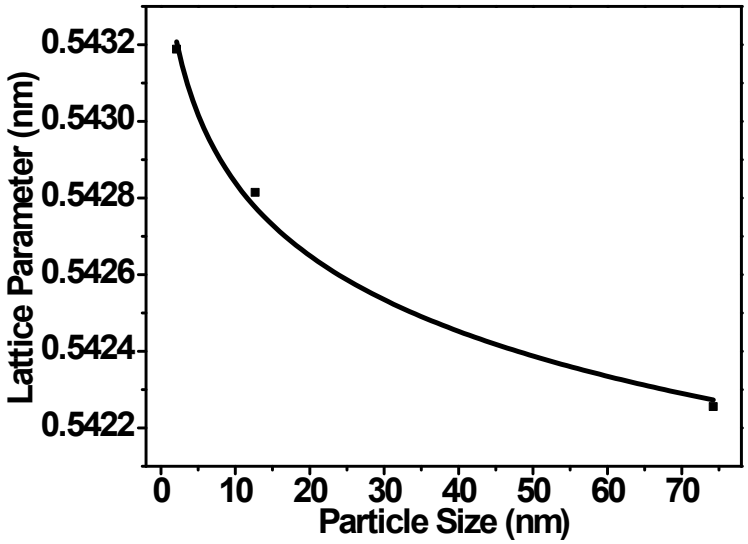
Figure 3. TEM images (a and b) and corresponding particle size distribution (c and d) of S2 sample.

Figure 4. Forecast with PLSR of cell viability at different mean particles size (a) S1-2.1 nm, S2-6.8 nm, S3 73.9 nm, (b) S1-2.1 nm, S2-12.7 nm, S3-73.9 nm, (c) S1-2.1 nm, S2-15.2 nm, S3 73.9 nm, and (d) S1-2.1 nm, S2-32.8 nm, S3 73.9 nm (▲represents S1 sample●represents S2 sample ◆represents S3 sample)

Figure 5. Particle size distribution of S1, S2 and S3 samples in ultrapure water (a-c) and in cell culture medium (d-f) at 200 ng/uL concentration determined by dynamic light scattering.

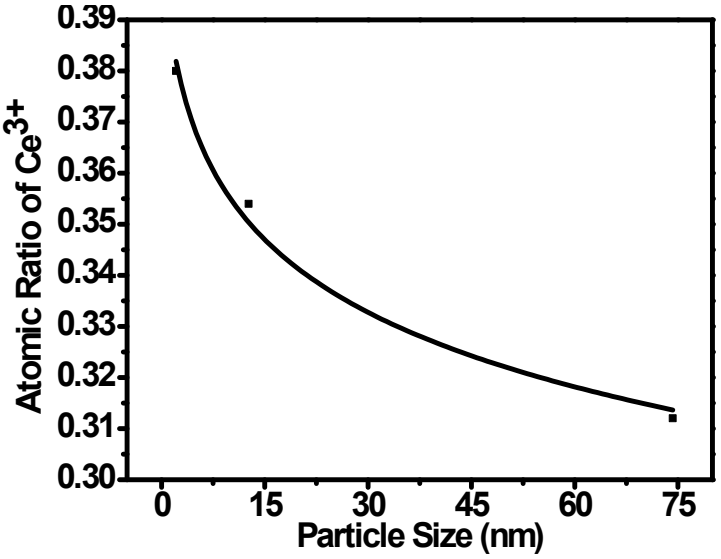
Figure 6. The relationship between the percent of variance explained in the response variable and the number of components.

Figure 1.



$$y = 0.54 - 2.96 \times 10^{-4} ln(x + 1.071)$$

Figure 2.



$$y = 0.4069 - 0.02158 * ln(x + 1.0706)$$

Figure 3.

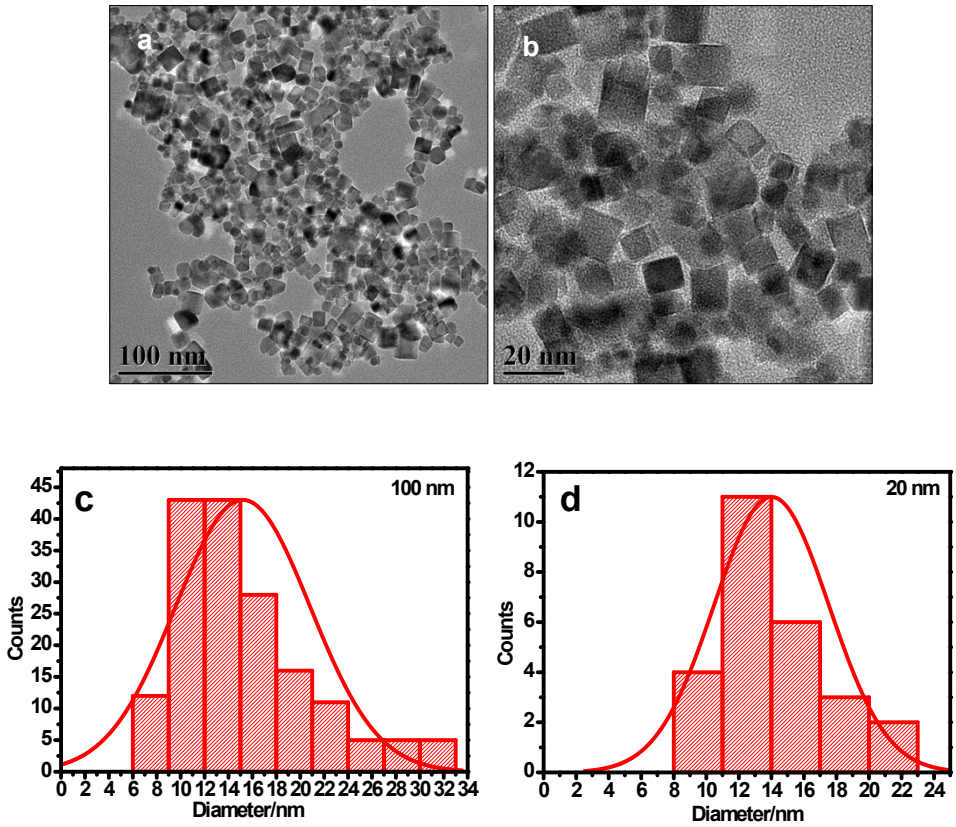
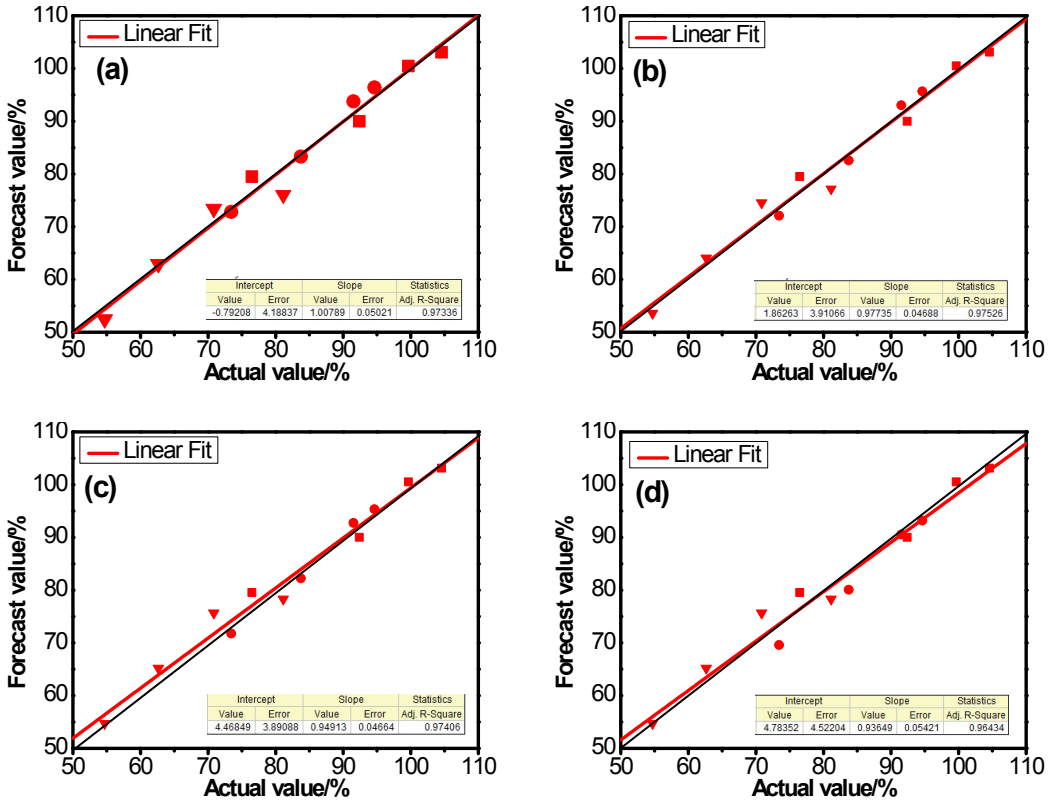

a
100 nm
b
20 nm
c
100 nm
Counts
Diameter/nm
d
20 nm
Counts
Diameter/nm

Figure 4.



The particle size acted as a significant factor in their regression models, changes in particle size could cause the potential impacts on their resultant findings. PLSR is more reliable for adjusting a model for output prediction[1]. It is worth noting that partial least squares regression model has some tolerance when the particle size of S2 sample deviate from the average. In order to demonstrate the impacts to this partial least square regression model, we added a prediction of cell viability with different particle sizes in the Support Information. We input four groups of particle sizes (S1-2.1 nm, S2-6.8 nm, S3-73.9 nm; S1-2.1 nm, S2-12.7 nm, S3-73.9 nm; S1-2.1 nm, S2-15.2 nm, S3 73.9 nm and S1-2.1 nm, S2-32.8 nm, S3 73.9 nm) and then calculate using the regression model (Figure S4). The diagonal line indicates that the theoretical prediction is equal to the experimental test value. The red line represents the fitted line of the regression model when the particle size of S2 sample is changed in each group. From the forecast value of regression model, the R-square of the fit are 0.97336, 0.97526, 0.97406 and 0.96434, respectively. The percent errors between the calculation and the experiment are 2.41%, 2.43%, 2.45%, 2.95%, respectively. It is clearly exhibited that there is a 2.95% deviation between the forecast value and actual experiment value when the particle size of S2 sample was 32.8 nm [2]. It also indicated that 32.8 nm

particle size should be a significant factor in their regression model. However, when the percent error between prediction and experiment exceed a certain threshold (⩾5%), it is better to reconsider the regression method, to reexamine the input data, to screened outliers and to review the internal correlation between independent and dependent variables.

Figure 5.

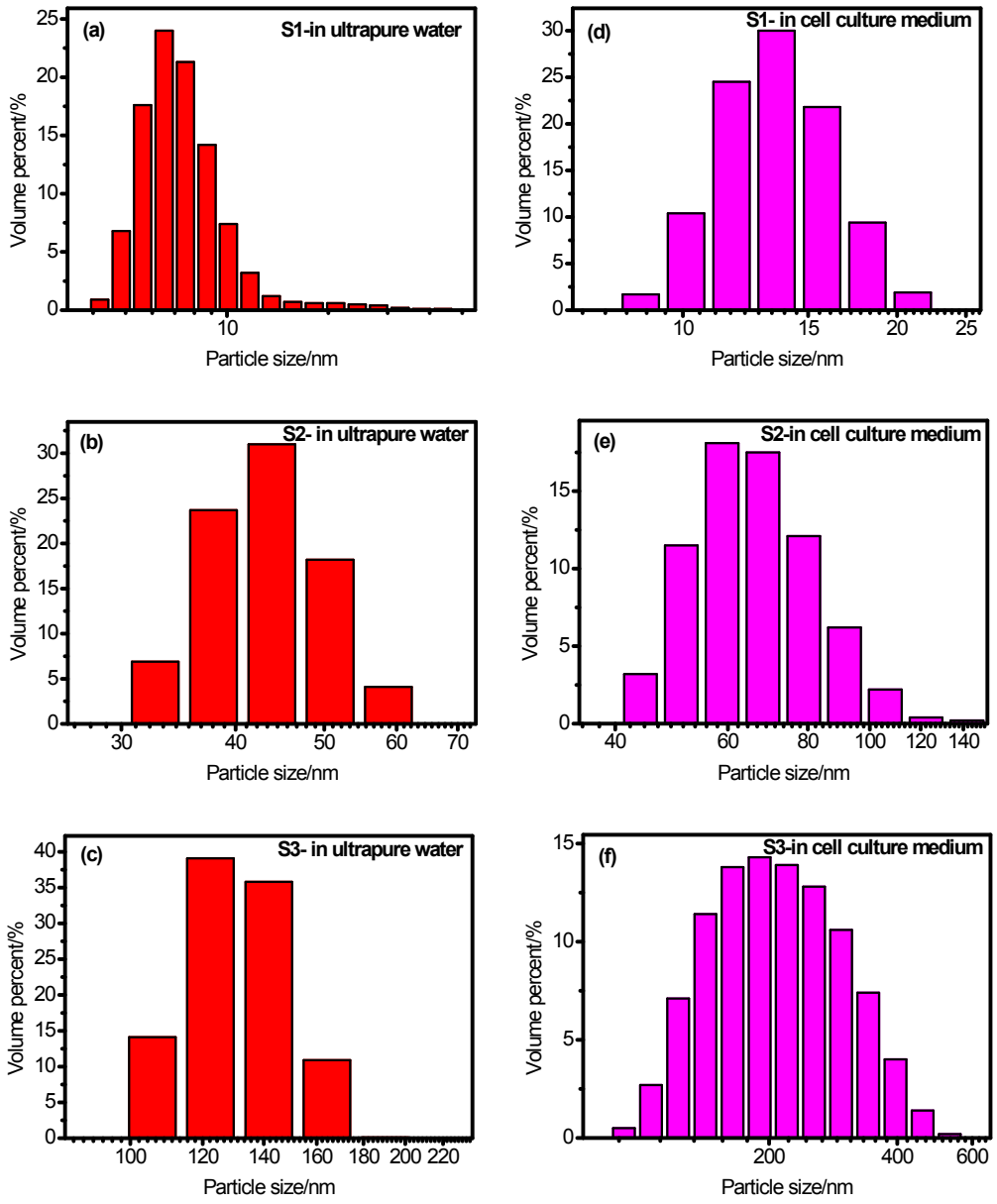
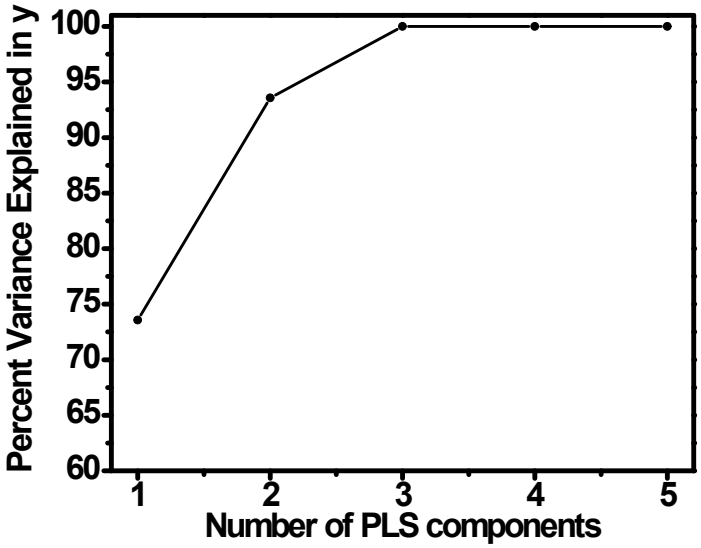


Figure S5 showed that the particle size distribution was larger than those of TEM images. These phenomena exhibited that Ce NPs agglomerate in ultrapure water and in cell culture medium. These results were mainly caused by the change of electronic repulsive force and the electrostatic attraction. In ultrapure water, electrostatic repulsive forces had decreased between Ce NPs, which resulted in Ce NPs agglomerating. While in cell culture medium, negative proteins were adsorbed by electrostatic attraction to the positively charged S1 surface. As a result, the net interaction force represented by electrostatic repulsive forces and van der Waals attraction force was changed. Electrostatic attraction force became dominant against the electronic repulsive force and der Waals attraction force, and then lead to S1 sample agglomeration, which could also

explain the particle size increase of S2 and S3 samples were mixed with cell culture medium[3-4].

Figure 6.



The number of principal components was determined as 3 by 'pcacov' function in Matlab software. It was calculated the relationship between the percent of variance explained in the response variable and the number of principal components for further PLSR, as it was shown in Figure S6. From Figure S6, it could be observed that the ratio of the characteristic value is greater than 97.75% or even close 100% when the number of principal component was 3.

Table Captions

Table 1. Composition of RPM1 Cell Culture Medium 1640 [5]

Table 2. Comparison table between a partial least-square regression (PLSR) method and Matlab code

Table 1.

Constituent	Concentration, mg/liter
Amino acids	
l-Arginine, **positive charge**	200
l-Asparagine	50
l-Aspartic acid, **negative charge**	20
l-Cystine	50
l-Glutamic acid, **negative charge**	20
l-Glutamine	300
Glutathione, reduced	1
Glycine	10
l-Histidine, **positive charge**	15
l-Hydroxyproline	20
l-Isoleucine, **positive charge**	50
l-Leucine, **positive charge**	50
l-Lysine hydrochloride	40
l-Methionine, **positive charge**	15
l-Phenylalanine	15
l-Proline	20
l-Serine	30
l-Threonine	20
l-Tryptophan	5
l-Tyrosine, positive charge	20
l-Valine	20
Vitamins	
para-Aminobenzoic acid	1
Biotin	0.2
Calcium pantothenate	0.25
Choline chloride	3
Cyanocobalamin	0.005
Folic acid	1
l-Inositol	35
Nicotinamide	1
Pyridoxine hydrochloride	1
Riboflavin	0.2
Thiamine hydrochloride	1
Salts	
Calcium nitrate tetrahydrate	100
Disodium phosphate heptahydrate	1512
Magnesium sulfate heptahydrate	100
Potassium chloride	400
Sodium bicarbonate	2000
Sodium chloride	6000
Miscellaneous	
Glucose	2000
Phenolsulfonphthalein	5

Table 2.

Step 1	Clear all input and output from the Command Window display, and remove items from workspace, freeing up system memory.	`clc,clear`

Step 2	Create the array of cytotoxicity according to Table 1.
	
	The first to fifth column represent the five independent variables respectively. From left to right: CeNPs particle size, BET surface area, $[Ce^{3+}]$, zeta-potential in cell culture medium and concentration.
	
	The sixth column represents the dependent variable: cell viability.
		`CeNPs = [2.12 134.2371 0.38 -6.23 20 104.56`
		`2.12 134.2371 0.38 -6.23 40 99.65`
		`2.12 134.2371 0.38 -6.23 120 92.39`
		`2.12 134.2371 0.38 -6.23 200 76.50`
		`12.66 117.3064 0.354 -6.75 20 94.63`
		`12.66 117.3064 0.354 -6.75 40 91.53`
		`12.66 117.3064 0.354 -6.75 120 83.73`
		`12.66 117.3064 0.354 -6.75 200 73.44`
		`74.23 114.1727 0.312 -8.26 20 81.14`
		`74.23 114.1727 0.312 -8.26 40 70.85`
		`74.23 114.1727 0.312 -8.26 120 62.66`
		`74.23 114.1727 0.312 -8.26 200 54.70];`

	Standardize the array.	`zsCeNPs = zscore(CeNPs);`
	Divide the independent variable and dependent variable into two matrixes, *XX* and *YY* respectively.	`XX = zsCeNPs(:,1:end-1);` `YY = zsCeNPs(:,end);`
Step 3	Calculate the covariance matrix of the array of cytotoxicity as function 'pcacov' requires covariance matrix input.	`r = cov(CeNPs);`

	Return a vector '*rate*' containing the percentage of the total variance explained by each principal component.	`[vec1, lamda, rate] = pcacov(XX);`

	Plot the percent of variance explained in the response variable as a function of the number of components.	`contr = cumsum(rate)';`
		`plot(1:6,contr,'-bo');`
		`ylim([0 100]);`
		`xlabel('Number of PLS components'); ylabel('Percent Variance Explained in y');`

	According to the above plot, an appropriate principal number of components can be confirmed and then be input as *ncomp*.	`ncomp = input('The number of components = ')`
Step 4	Computes a partial least-squares (PLS) regression of dependent variable on independent variable, using *ncomp* principal components, and returns the PLS regression coefficients *BETA*.	`[XL,YL,XS,YS,BETA] = plsregress(XX,YY,ncomp);`

	Calculate the constant term of regression equation	`n = size(XX,2);`
		`mu = mean(CeNPs);`
		`sig = std(CeNPs);`
		`beta2(1) = mu(end)-mu(1:n)./sig(1:n)*BETA([2:end]).*sig(end);`
		`beta2(2:n+1) = (1./sig(1:n))'*sig(n+1:end).*BETA([2:end])`

Step 5	Use the regression equation to get a series of model value and make a plot of model value as a function of actual value to test the reliability of regression equation.	`yfit = beta2(1)+beta2(2)*CeNPs(:,1)+beta2(3)*CeNPs(:,2)+beta2(4)*CeNPs(:,3)+beta2(5)*CeNPs(:,4)+beta2(6)*CeNPs(:,5);`
		`plot(CeNPs(:,6)',yfit','o');`
		`xlabel('Actual value');`
		`ylabel('Model value')`

	Test the reliability of regression equation by assessing whether the array of model and actual value have close population mean. The returned h value is 0 indicates that at the 1% significance level, regression equation is reliable.	`[h,p] = ttest(CeNPs(:,6),yfit,0.01);`
	Furthermore, we introduce percent error to quantitatively measure the closeness of the fit between the model and actual data.	`D = mean(abs((yfit - CeNPs(:,6))./CeNPs(:,6)))*100`

```
clc,clear
CeNPs = [2.12 134.2371 0.38 -6.23 20 104.56
2.12 134.2371 0.38 -6.23 40 99.65
2.12 134.2371 0.38 -6.23 120 92.39
2.12 134.2371 0.38 -6.23 200 76.50
12.66 117.3064 0.354 -6.75 20 94.63
12.66 117.3064 0.354 -6.75 40 91.53
12.66 117.3064 0.354 -6.75 120 83.73
12.66 117.3064 0.354 -6.75 200 73.44
74.23 114.1727 0.312 -8.26 20 81.14
74.23 114.1727 0.312 -8.26 40 70.85
74.23 114.1727 0.312 -8.26 120 62.66
74.23 114.1727 0.312 -8.26 200 54.70];
zsCeNPs = zscore(CeNPs);
XX = zsCeNPs(:,1:end-1); YY = zsCeNPs(:,end);
r = cov(CeNPs);
[vec1, lamda, rate] = pcacov(r);
contr = cumsum(rate)';
plot(1:6,contr,'-bo');ylim([0 100]);
xlabel('Number of PLS components'); ylabel('Percent Variance
Explained in y');
ncomp = input('The number of components = ')
[XL,YL,XS,YS,BETA] = plsregress(XX,YY,ncomp);
n = size(XX,2);mu = mean(CeNPs);sig = std(CeNPs);
beta2(1) = mu(end)-mu(1:n)./sig(1:n)*BETA([2:end]).*sig(end);
beta2(2:n+1) = (1./sig(1:n))'*sig(n+1:end).*BETA([2:end]);
yfit =
```

```
beta2(1)+beta2(2)*CeNPs(:,1)+beta2(3)*CeNPs(:,2)+beta2(4)*CeNPs(:,3)+
beta2(5)*CeNPs(:,4)+beta2(6)*CeNPs(:,5);
plot(CeNPs(:,6)',yfit','o');xlabel('Actual value');ylabel('Model
value')
[h,p] = ttest(CeNPs(:,6),yfit,0.01)
D = mean(abs((yfit - CeNPs(:,6))./CeNPs(:,6)))*100
```

Reference

1. J. L. Godoy, J. R. Vega and J. L. Marchetti, *Chemom. Intell. Lab. Syst.*, 2014, **130**, 182–191.
2. H. Nawaz, F. Bonnier, P. Knief, O. Howe, F. M. Lyng, A. D. Meade and H. J. Byrne, *Analyst*, 2010, **135**, 3070-3076.
3. D. Kim, H. El-Shall, D. Dennis and T. Morey, *Colloids Surf.,B*, 2005, **40**, 83–91.
4. K. Rezwan, L. P. Meier, M. Rezwan, J. Vörös, M. Textor and L. J. Gauckler, *Langmuir*, 2004, **20**, 10055–10061.
5. G. E. Moore, R. E. Gerner and H. A. Franklin, *Jama*, 1967, **199**, 519–524.