Supplementary file

Efforts to untie the multicollinearity knot and identify factors controlling macropore structures in shale oil reservoirs

Ziyi Wang[1], Lin Dong[1,*], Zhijun Jin[1,2], Shuangmei Zou[3], Jinhua Fu[4], Rukai Zhu[5]

[1] *School of Earth and Space Sciences, Peking University, Beijing 100871, P. R. China*

[2] *Institute of Energy, Peking University, Beijing 100871, P. R. China*

[3] *School of Earth Resources, China University of Geosciences (Wuhan), Wuhan 430074, P. R. China*

[4] *PetroChina Changqing Oilfield Company, Xi'an 710018, P. R. China*

[5]*Research Institute of Petroleum Exploration & Development, PetroChina, Beijing 100083, P. R. China*

E-mail address: ziyi-wang@pku.edu.cn (Z. Wang); lin.dong@pku.edu.cn (L. Dong); jinzj1957@pku.edu.cn (Z. Jin); zousm@cug.edu.cn (S. Zou); fjh_cq@petrochina.com.cn (J. Fu); zrk@petrochina.com.cn (R. Zhu).

[*] Corresponding author (ORCID: 0000-0003-1754-842X)

Wang, Z., Dong, L., Jin, Z., et al. Efforts to untie the multicollinearity knot and identify factors controlling macropore structures in shale oil reservoirs. Advances in Geo-Energy Research, 2024, 11(3): 194-207.

The link to this file is: https://doi.org/10.46690/ager.2024.03.04

This file includes:

Supplementary Notes

Detailed method of the partial least square (PLS) regression analysis

Partial least squares (PLS) regression addresses the limitations inherent in traditional multiple regression methods when dealing with multicollinear variables and has garnered extensive application in correlation analyses between pore properties and geological factors (Liu et al., 2017, 2019a, 2019b). This PLS regression technique enables regression analysis between a set (or multiple sets) of interrelated independent and dependent variables. Within the domain of shale pore research, the prevalent approach involves examining the correlations between a multitude of geological factors and a singular pore-related parameter, with the detailed computational methodologies (Rännar et al., 1995; Stocchero et al., 2019; Wold et al., 2001) elaborated as follow:

Assuming that a partial least squares analysis is conducted to examine the correlation between one dependent variable and p independent variables across n samples. The independent variables are represented by x, the data for the i-th sample of the j-th independent variable is denoted by x_{ij}, and the sets of independent variables can be represented as an $n \times p$ matrix, which is denoted as $\boldsymbol{X}$. Correspondingly, the dependent variable is symbolized as y, the data for the dependent variable of the i-th sample is expressed as y_i, and the dependent variable set can be represented as a column vector, denoted as $\boldsymbol{y}$. Consequently, $\boldsymbol{X}$ and $\boldsymbol{y}$ can be respectively represented as:

$$\boldsymbol{X} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \tag{S1}$$

$$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \tag{S2}$$

The computation of PLS regression entails three pivotal stages. Step 1: all data are centralized and standardized via dividing each variable's data by its mean value, followed by division by its standard deviation. The standardized values of x_{ij} are denoted as x_{ij}^*, while the standardized values of y_i are denoted as y_i^*:

$$x_{ij}^{*} = \frac{x_{ij} - \bar{x}_j}{\sigma_j} \tag{S3}$$

$$y_i^{*} = \frac{y_i - \bar{y}}{\sigma_y} \tag{S4}$$

where $\bar{x}_j$ and $\bar{y}$ represent the mean values of the j-th independent variable $\boldsymbol{x_j}$ and the dependent variable $\boldsymbol{y}$, respectively. Additionally, σ_j and σ_y denote the standard deviations of $\boldsymbol{x_j}$ and $\boldsymbol{y}$, respectively. The standardized dataset obtained by normalizing $\boldsymbol{X}$ is denoted as matrix $\boldsymbol{E_0}$. The set of standardized data for the j-th independent variable is denoted as $\boldsymbol{x_j^{*}}$, and the set of standardized data for $\boldsymbol{y}$ is represented by vector $\boldsymbol{F_0}$. After standardization, the measurement units of each variable are consistent.

Step 2: several orthogonal components are extracted from the variable sets. The first component, denoted as $\boldsymbol{t_1}$, is extracted from $\boldsymbol{E_0}$ by multiplying $\boldsymbol{E_0}$ with a weight vector:

$$\boldsymbol{t_1} = \boldsymbol{E_0 w_1} \tag{S5}$$

where $\boldsymbol{w_1}$ represents the weight vector, which is a unit vector of $\boldsymbol{E_0}$ known as the first axis of $\boldsymbol{E_0}$. The primary principle in extracting the $\boldsymbol{t_1}$ is to capture the maximum amount of variation present in $\boldsymbol{X}$, while simultaneously providing a maximized explanatory ability for $\boldsymbol{F_0}$. This process can be represented as:

$$\begin{cases} \mathrm{Var}(\boldsymbol{t_1}) \rightarrow \max \\ \mathrm{r}(\boldsymbol{t_1}, \boldsymbol{F_0}) \rightarrow \max \end{cases} \tag{S6}$$

where $\mathrm{Var}(\cdot)$ represents the variance operator, and $\mathrm{r}(\cdot,\cdot)$ denotes the correlation coefficient operator. This problem can be formulated as solving the following optimization problem:

$$\begin{cases} \max\langle \boldsymbol{E_0 w_1}, \boldsymbol{F_0} \rangle \\ \text{s.t. } \boldsymbol{w_1^{\mathrm{T}} w_1} = 1 \end{cases} \tag{S7}$$

That is to say that find the maximum value of $\boldsymbol{w_1^{\mathrm{T}} E_0^{\mathrm{T}} F_0}$ under the constraint condition of $\|\boldsymbol{w_1}\| = 1$. By optimizing calculations, the following regression equation can be obtained:

$$\boldsymbol{E_0} = \boldsymbol{t_1 p_1^{\mathrm{T}}} + \boldsymbol{E_1} \tag{S8}$$

$$\boldsymbol{F_0} = \boldsymbol{t_1} r_1 + \boldsymbol{F_1} \tag{S9}$$

where $\boldsymbol{p_1}$ and r_1 represent the coefficients (r_1 is a scalar), and $\boldsymbol{E_1}$ and $\boldsymbol{F_1}$ represent the residual matrices.

Step 3: Replace $\boldsymbol{E_0}$ and $\boldsymbol{F_0}$ with $\boldsymbol{E_1}$ and $\boldsymbol{F_1}$ respectively, and repeat the Step 2 in the same manner. Extracting the second component $\boldsymbol{t_2}$ through optimization calculation and obtain the following regression equation:

$$\boldsymbol{E_1} = \boldsymbol{t_2}\boldsymbol{p_2}^{\mathrm{T}} + \boldsymbol{E_2} \tag{S10}$$

$$\boldsymbol{F_1} = \boldsymbol{t_2} r_2 + \boldsymbol{F_2} \tag{S11}$$

where $\boldsymbol{p_2}$ and r_2 represent the coefficients (r_2 is a scalar), and $\boldsymbol{E_2}$ and $\boldsymbol{F_2}$ represent the residual matrices. Replace $\boldsymbol{E_1}$ and $\boldsymbol{F_1}$ with $\boldsymbol{E_2}$ and $\boldsymbol{F_2}$ respectively, and repeat the Step 3. Iterate this process (the number of iterations is determined through cross-validation, as described below) to obtain a series of regression equations.

If the rank of $\boldsymbol{X}$ is A, the regression equation for $\boldsymbol{F_0}$ is expressed as:

$$\boldsymbol{F_0} = \boldsymbol{t_1} r_1 + \boldsymbol{t_2} r_2 + \cdots + \boldsymbol{t_A} r_A + \boldsymbol{F_A} \tag{S12}$$

where $\boldsymbol{F_A}$ is a residual matrix.

As $\boldsymbol{t_1}$, $\boldsymbol{t_2}$, $\cdots$, $\boldsymbol{t_A}$ are all linear equations of $\boldsymbol{x_1^*}$, $\boldsymbol{x_2^*}$, $\cdots$, $\boldsymbol{x_j^*}$, Eq. S12 can be transformed into a linear equation of the standardized dependent variable $\boldsymbol{y}^*$ (i.e., $\boldsymbol{F_0}$) with respect to the standardized independent variable $\boldsymbol{x_j^*}$:

$$\boldsymbol{y}^* = a_1 \boldsymbol{x_1^*} + a_2 \boldsymbol{x_2^*} + \cdots + a_p \boldsymbol{x_p^*} + \boldsymbol{F_A} \tag{S13}$$

where a_1, a_2, $\cdots$, a_m are constants and $\boldsymbol{F_A}$ is a residual matrix.

By performing inverse standardization on Eq. S13, the equation of $\boldsymbol{y}$ with respect to $\boldsymbol{x_j}$ can be obtained:

$$\boldsymbol{y} = b_1 \boldsymbol{x_1} + b_2 \boldsymbol{x_2} + \cdots + b_p \boldsymbol{x_p} + \boldsymbol{F_A} \tag{S14}$$

where b_1, b_2, $\cdots$, b_m are constants and $\boldsymbol{F_A}$ is a residual matrix.

When utilizing partial least squares regression to tackle practical problems, it is not necessary to include every component in the construction of the regression model. Instead, it is advantageous to

select only the first m components that contribute to a regression model with superior predictive performance. If subsequent components fail to provide additional meaningful information towards explaining the dependent variable, including too many components can lead to misinterpretation of statistical trends and result in inaccurate predictions. Hence, it is crucial to extract an appropriate number of components (m) for constructing the regression model, with m determined through cross-validation:

The dataset comprising n samples is divided into two distinct groups: the first group consists of the remaining samples ($n-1$ samples) after the exclusion of the i-th sample. These samples are used to establish a regression equation involving h components. The second group comprises the sample which was initially excluded prior to the regression. This particular sample's data is then inserted into the established regression equation, enabling the derivation of the regression value $\hat{y}_{h(-i)}$ for y_i with respect to the i-th sample. By applying the aforementioned steps to all samples ($i = 1, 2, \cdots, n$), the predicted error sum of squares $S_{PRESS,h}$ for $\boldsymbol{y}$ can be calculated:

$$S_{PRESS,h} = \sum_{i=1}^{q} \left(y_i - \hat{y}_{h(-i)}\right)^2 \tag{S15}$$

A regression equation with poorer robustness and larger errors tends to have a relatively higher value of $S_{PRESS,h}$. Subsequently, by utilizing the entire sample set, a partial least squares regression equation is constructed with $h-1$ components. We designate $\hat{y}_{(h-1)i}$ as the regression value for the i-th sample. The sum of squared fitting errors of $\boldsymbol{y}$, denoted as $S_{SS,h-1}$, is computed:

$$S_{SS,h-1} = \sum_{i=1}^{q} \left(y_i - \hat{y}_{(h-1)i}\right)^2 \tag{S16}$$

In a PLS model, a lower $\frac{S_{PRESS,h}}{S_{SS,h-1}}$ indicates better performance. Typically, when $\frac{S_{PRESS,h}}{S_{SS,h-1}} \leq 0.95^2$, incorporating the h-th component $\boldsymbol{t_h}$ can substantially enhance the predictive accuracy of the regression model. Hence, within the confines of $\frac{S_{PRESS,h}}{S_{SS,h-1}} \leq 0.95^2$, the maximum value of h signifies the optimal number of extracted components.

It is crucial to acknowledge that in PLS analysis, the use of standardized coefficients (the coefficients in regression equation constructed from standardized data) might not accurately portray the extent of influence (i.e. marginal contribution) of descriptors on responses, partially due to potential intercorrelations among the variables. To address this issue, a parameter termed variable importance in projection (VIP) is employed in this study. VIP quantifies the incremental contribution of a descriptor to the extracted component(s) (Wang et al., 2006; Favilla et al., 2013). It is calculated using the following equation:

$$\mathrm{VIP}(x_j) = \sqrt{\frac{p}{\mathrm{Rd}(\boldsymbol{y};\ \boldsymbol{t_1}, \boldsymbol{t_2}, \cdots, \boldsymbol{t_m}\ )} \sum_{h=1}^{m} \mathrm{Rd}(\boldsymbol{y}; \boldsymbol{t_h}) \omega_{hj}^2} \tag{S17}$$

where $\mathrm{VIP}(x_j)$ represents the VIP value of the independent variable $\boldsymbol{x_j}$. $\mathrm{Rd}(\boldsymbol{y}; \boldsymbol{t_h})$ denotes the variation precision of $\boldsymbol{y}$ explained by $\boldsymbol{t_h}$ (i.e., the proportion of $\boldsymbol{y}$'s total variation accounted for by the variation influenced by $\boldsymbol{t_h}$), which is denoted as $\mathrm{Rd}(\boldsymbol{y}; \boldsymbol{t_h}) = \mathrm{r}^2(\boldsymbol{y}; \boldsymbol{t_h})$. $\mathrm{Rd}(\boldsymbol{y};\ \boldsymbol{t_1}, \boldsymbol{t_2}, \cdots, \boldsymbol{t_m}\ )$ represents the cumulative variation precision of $\boldsymbol{y}$ explained by $\boldsymbol{t_1}, \boldsymbol{t_2}, \cdots, \boldsymbol{t_m}$. It is calculated by summing the individual variation precisions, denoted as $\mathrm{Rd}(\boldsymbol{y};\ \boldsymbol{t_1}, \boldsymbol{t_2}, \cdots, \boldsymbol{t_m}\ ) = \sum_{h=1}^{m} \mathrm{Rd}(\boldsymbol{y};\ \boldsymbol{t_h})$. ω_{hj} denotes the weight of $\boldsymbol{x_j}$ on the h-th axis ($\boldsymbol{w_h}$). When a descriptor plays an important role in controlling the component(s), it significantly contributes to elucidating the response and is characterized by a relatively high VIP value. Generally, descriptors with VIP values greater than 1 (the average of square VIP values) are considered pertinent and vital for predicting the response (Jia et al., 2009; Favilla et al., 2013; Stocchero et al., 2019).

Supplementary Figures

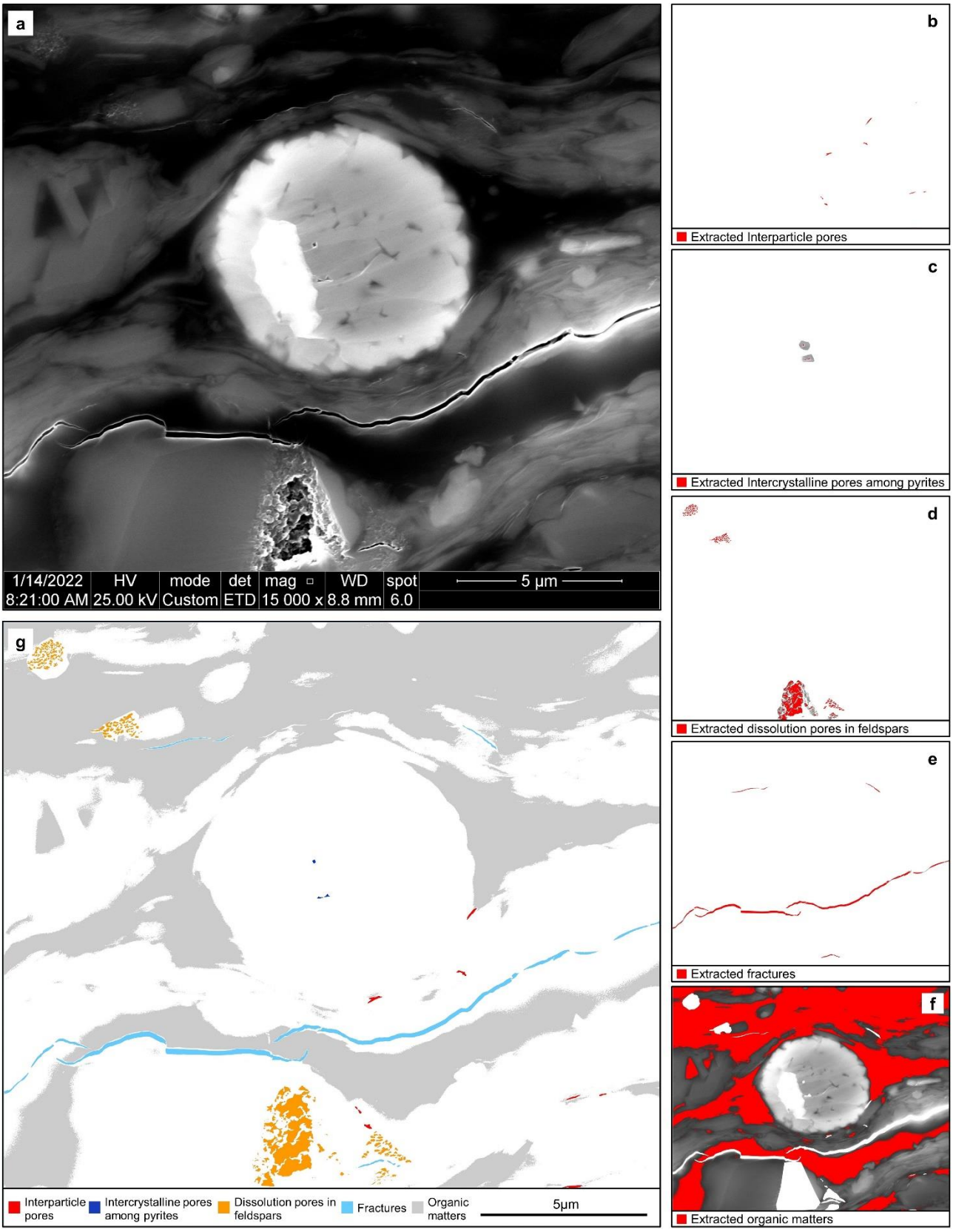

Fig. S1 Process of SEM quantitative analysis. (**a**) A secondary electron image of sample #15. (**b**–**f**) Pore and organic matter extraction from the lassoed SEM images separately containing interparticle pores (**b**), intercrystalline pores of pyrite (**c**), dissolution pores of feldspar (**d**), fractures (**e**) and organic matters (**f**) by using Fiji-imageJ software. (**g**) Extracted organic matters and pores with different genetic types.

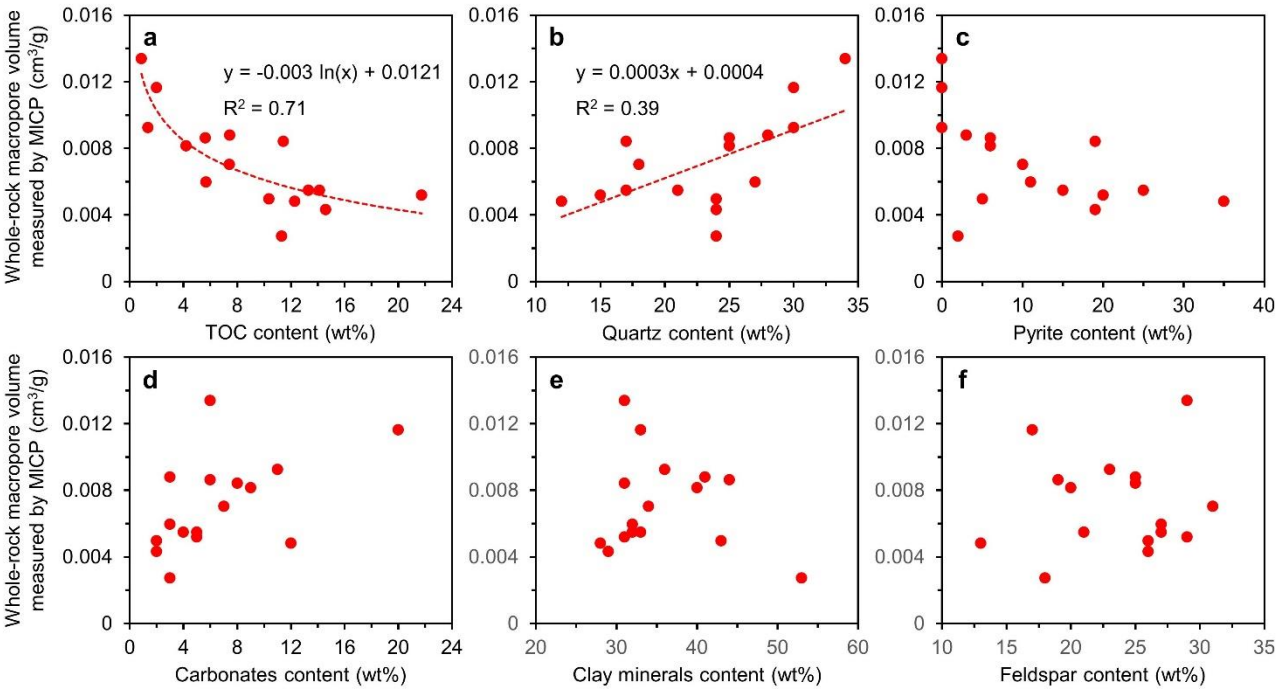

Fig. S2 Relationships between whole-rock macropore volume measured by mercury injection capillary pressure (MICP) and geological factors. (**a**) Macropore volume shows an evident logarithmic relationship with TOC content. (**b**) Macropore volume shows a linear relationship with TOC content. (**c**–**f**) Macropore volume shows ambiguous correlations with pyrite (**c**), carbonates (**d**), clay minerals (**e**) and feldspar (**f**).

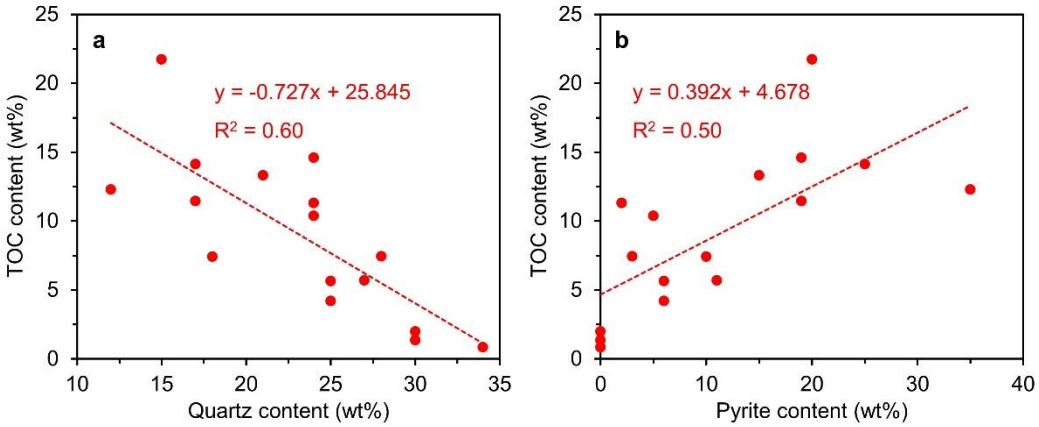

Fig. S3 Relationships of TOC with quartz and pyrite. (**a**) TOC content has an obvious negative correlation with quartz content. (**b**) TOC content has an obvious positive correlation with pyrite content.

Table S1 Sampling information, lithology, total organic carbon (TOC), vitrinite reflectance (R_o) and mineral composition of the Chang-7 shale oil reservoir samples of the Ordos Basin.

Sample No.	Well	Depth (m)	Lithology	R_o (%)	TOC (wt%)	Mineral content (wt%)				
						Clay minerals	Quartz	Feldspar	Carbonates	Pyrite
#1	YY1	191.2	MS	0.66	1.35	36	30	23	11	n.d.
#2	YY1	209.7	MS	0.63	1.99	33	30	17	20	n.d.
#3	YY1	247.7	MS	0.69	0.85	31	34	29	6	n.d.
#4	W336	1963.8	MS	0.73	7.44	41	28	25	3	3
#5	W336	2059.5	MS	0.77	10.38	43	24	26	2	5
#6	W336	2020.7	CS	0.73	11.31	53	24	18	3	2
#7	YY1	224.0	CS+SL	0.66	11.45	31	17	25	8	19
#8	YY1	225.8	CS+SL+MS	0.66	14.59	29	24	26	2	19
#9	YY1	228.4	CS+SL	0.65	5.64	44	25	19	6	6
#10	YY1	231.8	CS+SL	0.70	7.43	34	18	31	7	10
#11	YY1	232.1	CS+SL+MS	0.66	4.20	40	25	20	9	6
#12	YY1	233.2	CS+SL+MS	0.68	13.31	32	21	27	5	15
#13	B522	1949.3	CS+SL	0.89	12.29	28	12	13	12	35
#14	B522	1944.4	CS+SL	0.92	5.69	32	27	27	3	11
#15	B522	1940.5	CS+SL	0.96	14.13	33	17	21	4	25
#16	W336	1955.4	CS+SL	0.73	21.73	31	15	29	5	20
Average				0.73	8.99	35.7	23.2	23.5	6.6	11.0

MS: massive siltstone; CS: clay shale; SL: silty lamina; n.d.: no detected (below detection limits).

Table S2 Pore parameters obtained by mercury injection capillary pressure (MICP) and scanning electron microscope (SEM) quantitative analysis as well as area ratio of organic matters obtained by SEM for the whole-rock samples from the Chang-7 shale oil reservoirs of the Ordos Basin.

Sample No.	Lithology	Macropore volume obtained by MICP (cm^3/g)	Total surface macroporosity obtained by SEM (%)	Surface macroporosity of different pore types obtained by SEM (%)						Area ratio of organic matters obtained by SEM (%)
				Pores among clay minerals	Interparticle pores	Dissolution pores in feldspar	Dissolution pores in carbonates	Fractures	Other pores	
#1	MS	0.0092	1.71	1.50	0.11	0.05	n.o.	0.04	0.01	4.12
#2	MS	0.0116	2.80	1.97	0.35	0.13	0.23	0.09	0.02	1.46
#3	MS	0.0134	3.22	2.83	0.25	0.01	n.o.	0.10	0.03	0.72
#4	MS	0.0088	2.37	1.38	0.53	0.39	n.o.	0.06	0.00	6.40
#5	MS	0.0050	1.38	1.11	0.18	0.04	n.o.	0.05	0.00	8.95
#6	CS	0.0027	0.71	0.37	0.06	0.26	n.o.	0.00	0.01	6.78
#7	CS+SL	0.0084	1.24	0.71	0.10	0.15	0.13	0.12	0.03	15.09
#8	CS+SL+MS	0.0043	0.74	0.32	0.19	0.15	n.o.	0.05	0.03	13.01
#9	CS+SL	0.0086	1.41	0.96	0.08	0.32	0.01	0.03	0.01	8.00
#10	CS+SL	0.0070	1.17	0.88	0.19	0.09	n.o.	0.01	0.00	10.57
#11	CS+SL+MS	0.0082	1.56	1.08	0.31	0.12	0.00	0.04	0.01	6.82
#12	CS+SL+MS	0.0055	0.93	0.72	0.09	0.07	n.o.	0.04	0.01	10.00
#13	CS+SL	0.0048	0.90	0.48	0.17	0.13	0.04	0.02	0.06	8.75
#14	CS+SL	0.0060	1.34	0.99	0.12	0.18	0.00	0.03	0.01	6.25
#15	CS+SL	0.0055	0.59	0.33	0.04	0.05	0.01	0.14	0.02	17.82
#16	CS+SL	0.0052	0.77	0.71	0.03	0.01	n.o.	0.00	0.02	22.76
Average		0.0071	1.43	1.02	0.17	0.14	0.03	0.05	0.02	9.22

MS: massive siltstone; CS: clay shale; SL: silty lamina; n.o.: no observed.

Table S3 Pore parameters and area ratio of organic matters obtained by scanning electron microscope (SEM) quantitative analysis for the clay shale within the Chang-7 shale oil reservoirs of the Ordos Basin.

Lithology No.	Sample No.	Lithology	Total surface macroporosity obtained by SEM (%)	Surface macroporosity of different pore types obtained by SEM (%)						Area ratio of organic matters obtained by SEM (%)
				Pores among clay minerals	Interparticle pores	Dissolution pores in feldspar	Dissolution pores in carbonates	Fractures	Other pores	
CS-6	#6	Clay shale	0.71	0.37	0.06	0.26	n.o.	0.00	0.01	6.78
CS-7	#7	Clay shale	0.49	0.28	0.01	0.02	n.o.	0.16	0.02	20.59
CS-8	#8	Clay shale	0.23	0.15	0.04	0.02	n.o.	0.01	0.01	21.04
CS-9	#9	Clay shale	0.97	0.72	0.09	0.10	0.01	0.03	0.01	9.58
CS-10	#10	Clay shale	0.46	0.34	0.10	0.01	n.o.	0.00	0.00	16.02
CS-11	#11	Clay shale	0.73	0.47	0.16	0.04	0.00	0.05	0.01	7.83
CS-12	#12	Clay shale	0.32	0.23	0.04	0.03	n.o.	0.00	0.01	22.05
CS-13	#13	Clay shale	0.60	0.28	0.12	0.10	n.o.	0.03	0.07	9.56
CS-14	#14	Clay shale	1.28	0.96	0.12	0.17	n.o.	0.02	0.01	6.31
CS-15	#15	Clay shale	0.60	0.34	0.04	0.05	0.01	0.15	0.02	17.81
CS-16	#16	Clay shale	0.73	0.67	0.03	0.01	n.o.	0.00	0.02	23.52
Average			0.65	0.44	0.07	0.07	0.00	0.04	0.02	14.64

n.o.: no observed.

Table S4 Pore parameters and area ratio of organic matters obtained by scanning electron microscope (SEM) quantitative analysis for the massive siltstone within the Chang-7 shale oil reservoirs of the Ordos Basin.

Lithology No.	Sample No.	Lithology	Total surface macroporosity obtained by SEM (%)	Surface macroporosity of different pore types obtained by SEM (%)						Area ratio of organic matters obtained by SEM (%)
				Pores among clay minerals	Interparticle pores	Dissolution pores in feldspar	Dissolution pores in carbonates	Fractures	Other pores	
MS-1	#1	Massive siltstone	1.71	1.50	0.11	0.05	n.o.	0.04	0.01	4.12
MS-2	#2	Massive siltstone	2.80	1.97	0.35	0.13	0.23	0.09	0.02	1.46
MS-3	#3	Massive siltstone	3.22	2.83	0.25	0.01	n.o.	0.10	0.03	0.72
MS-4	#4	Massive siltstone	2.37	1.38	0.53	0.39	n.o.	0.06	0.00	6.40
MS-5	#5	Massive siltstone	1.38	1.11	0.18	0.04	n.o.	0.05	0.00	8.95
MS-8	#8	Massive siltstone	1.40	0.54	0.39	0.32	n.o.	0.10	0.06	3.20
MS-11	#11	Massive siltstone	3.76	2.78	0.80	0.17	n.o.	0.01	0.00	4.11
MS-12	#12	Massive siltstone	1.31	1.06	0.12	0.06	n.o.	0.06	0.00	2.49
Average			2.24	1.65	0.34	0.15	0.03	0.06	0.02	3.93

n.o.: no observed.

Table S5 Pore parameters and area ratio of organic matters obtained by scanning electron microscope (SEM) quantitative analysis for the silty lamina within the Chang-7 shale oil reservoirs of the Ordos Basin.

Lithology No.	Sample No.	Lithology	Total surface macroporosity obtained by SEM (%)	Surface macroporosity of different pore types obtained by SEM (%)						Area ratio of organic matters obtained by SEM (%)
				Pores among clay minerals	Interparticle pores	Dissolution pores in feldspar	Dissolution pores in carbonates	Fractures	Other pores	
SL-7(1)	#7	Silty lamina	0.40	0.21	0.02	0.07	0.09	0.00	0.00	11.67
SL-7(2)	#7	Silty lamina	0.63	0.15	0.05	0.06	0.37	0.01	0.00	11.58
SL-7(3)	#7	Silty lamina	0.24	0.07	0.07	0.08	0.02	0.00	0.00	10.13
SL-7(4)	#7	Silty lamina	0.58	0.13	0.06	0.39	0.00	0.00	0.00	14.18
SL-7(5)	#7	Silty lamina	3.17	1.86	0.31	0.46	0.43	0.06	0.05	3.01
SL-8(1)	#8	Silty lamina	0.31	0.24	0.02	0.00	n.o.	0.04	0.00	15.09
SL-8(2)	#8	Silty lamina	0.23	0.17	0.02	0.02	n.o.	0.03	0.00	14.76
SL-9(1)	#9	Silty lamina	2.56	1.58	0.06	0.88	n.o.	0.03	0.00	3.86
SL-10(1)	#10	Silty lamina	2.08	1.57	0.30	0.18	n.o.	0.03	0.00	3.55
SL-11(1)	#11	Silty lamina	3.26	1.66	0.11	1.47	n.o.	0.02	0.00	5.08
SL-12(1)	#12	Silty lamina	0.40	0.35	0.04	0.00	n.o.	0.00	0.00	13.92
SL-12(2)	#12	Silty lamina	0.29	0.09	0.02	0.18	n.o.	0.00	0.00	9.77
SL-12(3)	#12	Silty lamina	1.21	0.68	0.05	0.43	n.o.	0.05	0.00	7.20
SL-13(1)	#13	Silty lamina	2.97	1.81	0.72	0.38	0.05	0.00	0.00	5.30
SL-13(2)	#13	Silty lamina	2.64	1.60	0.44	0.32	0.27	0.00	0.01	3.89
SL-14(1)	#14	Silty lamina	1.87	1.26	0.15	0.32	0.01	0.11	0.02	5.63
SL-15(1)	#15	Silty lamina	0.53	0.26	0.02	0.24	n.o.	0.01	0.00	18.87
SL-16(1)	#16	Silty lamina	1.37	1.27	0.02	0.06	n.o.	0.00	0.01	13.41
Average			1.37	0.83	0.14	0.31	0.07	0.02	0.01	9.50

n.o.: no observed.

References

Favilla, S., Durante, C., Vigni, M.L., et al. Assessing feature relevance in NPLS models by VIP. Chemometrics and Intelligent Laboratory Systems, 2013, 129: 76–86.

Jia, J., Deng, H., Duan, J., et al. Analysis of the major drivers of the ecological footprint using the STIRPAT model and the PLS method-A case study in Henan Province, China. Ecological Economics, 2009, 68: 2818–2824.

Liu, K., Ostadhassan, M., Sun, L., et al. A comprehensive pore structure study of the Bakken Shale with SANS, N2 adsorption and mercury intrusion. Fuel, 2019a, 245: 274–285.

Liu, K., Ostadhassan, M., Zhou, J., et al. Nanoscale pore structure characterization of the Bakken shale in the USA. Fuel, 2017, 209: 567–578.

Liu, K., Wang, L., Ostadhassan, M., et al. Nanopore structure comparison between shale oil and shale gas: examples from the Bakken and Longmaxi Formations. Petroleum Science, 2019b, 16: 77–93.

Rännar, S., Geladi, P., Lindgren, F., et al. A PLS kernel algorithm for data sets with many variables and few objects. Part II: Cross-validation, missing data and examples. Journal of Chemometrics, 1995, 9: 459–470.

Stocchero, M., Locci, E., D'Aloja, E., et al. PLS2 in metabolomics. Metabolites, 2019, 9: 9030051.

Wang, H., Wu, Z., Meng, J. Partial Least-Squares Regression–Linear and Nonlinear Methods. Beijing, China, National Defense Industry Press, 2006.

Wold, S., Sjostrom, M., Eriksson, L. PLS-regression : a basic tool of chemometrics. Chemometrics and Intelligent Laboratory Systems, 2001, 58: 109–130.