Pan, S., Horsfield, B., Zou, C., Yang, Z., Gao, D. (2017): Statistical analysis as a tool for assisting geochemical interpretation of the Upper Triassic Yanchang Formation, Ordos Basin, Central China. - *International Journal of Coal Geology*, *173*, pp. 51-64.

DOI: http://doi.org/10.1016/j.coal.2017.02.009

Statistical analysis as a tool for assisting geochemical interpretation of the Upper Triassic Yanchang Formation, Ordos Basin, Central China

Songqi Pan [a,b,c,*], Brian Horsfield [b], Caineng Zou [c], Zhi Yang [c,**], Dapeng Gao [d]

[a] *School of Earth and Space Sciences, Peking University, 100871 Beijing, China*

[b] *GFZ German Research Centre for Geosciences, 14473 Potsdam, Germany*

[c] *Research Institute of Petroleum Exploration and Development, PetroChina, 100083 Beijing, China*

[d] *Institute of Mechanics, Chinese Academy of Sciences, 100190 Beijing, China*

* Corresponding author: GFZ 3.2, Telegrafenberg, 14473 Potsdam, Germany.

E-mail address: pansongqi@pku.edu.cn (S.Q. Pan)

** Corresponding author: Department of petroleum geology, Xueyuan Road 20, 100083 Beijing, China.

E-mail address: yangzhi2009@petrochina.com.cn (Z. Yang)

ABSTRACT: The Yanchang Formation in the Ordos Basin is the most important petroleum play not only for conventional oil and gas accumulations, but also for newly emerging shale oil and tight gas resources. The molecular characterization of the basinwide source rocks predicts three groups of generative petroleum types: paraffinic high wax oil, mixed base (paraffinic-naphthenic-aromatic) low wax oil, and gas and condensate. Supplementary to previous work, 68 samples including the crude oils, source bitumens and reservoir extracts from the Yanchang petroleum play are analysed. The distribution of two terpane classes (eight tricyclic terpanes and eight pentacyclic terpanes) are determined with subsequent simultaneous RQ-mode factor analysis for a composite data set of these samples alongside 216 published crude oils worldwide with known facies descriptions. Thermal maturity has been evaluated as a consistent distribution at first using a combined method of a maturity-related biomarker [Ts/(Ts + Tm)] and aromatic parameters (Methyldibenzothiophene Ratios) to alleviate the maturity differences effect when discussing geochemical characterization. The R-mode factor analysis consists of the first two factors that are describing 45 present of the cumulative total variance in the data set, and presents a sample grouping pattern in Q-mode factor analysis which is determined by different contributions of terpane associations, i.e., the tricyclic C_{21} coupled with pentacyclic C_{26}, C_{27}, C_{28} and C_{30}, in the same factor space. Three terpane associations, the C_{26} and C_{28} terpanes, the C_{21} and C_{30} terpanes and the

C_{27} pentacyclic terpane, are respectively responsible for discriminating crude oil, reservoir extracts and source bitumens in RQ-mode factor analysis. Molecular compositions further address more detailed interrelationships among three sample groups that crude oils and reservoir extracts are sharing close genetic relationships both in depositional environment typing and C_{27}-C_{28}-C_{29} sterane distribution. Samples from source rocks vary much significantly. A mixing process which occurs after oils has been expelled from host source rocks into carrier units during accumulation. In addition, the migration-contamination of C_{29} sterols when oils are cross through the Chang 7-2 unit along migration pathways might also explain this lack of correlation between source rocks and oil-reservoir system.

Keywords: Upper Triassic Yanchang Formation; Ordos Basin; geochemistry; molecular composition; biomarker; simultaneous R- and Q-mode factor analysis;

1. Introduction

A source-oil correlation is defined as the discovery of a genetic relationship between a crude oil and its original source rock based on integrated geological and geochemical facts (Jones, 1987; Curiale, 1993). This relationship is of interest for scientific and economic reasons: it not only provides improved knowledge of the processes of petroleum generation, expulsion, migration and entrapment, but also improves the probability of commercial success based on a consideration of the petroleum fluids themselves rather than simply the reservoir, structure or trap in which the fluid is contained. Nevertheless, owing to the nature of "geological problems addressed with geochemical methods" (Curiale, 1993), the interdisciplinary character of source-oil correlation requires not merely detailed molecular and geochemical matching between oil and source rock, but also a consideration of what the circumstances were at the time the original source rock expelled a particular oil (Curiale, 1994).

The Mesozoic petroleum accumulations are a very important petroleum-bearing play in the Ordos Basin of Central China. Two Jurassic oil fields were firstly discovered during the 1970s followed soon thereafter by another seven (Qiu and Gong, 1999). In 1983, the Triassic Ansai oil field was discovered, and since then, more and more oil fields have been found in Triassic strata (Li and Lu, 2002). Over the past decades, the discovery of a series of oil accumulations in stratigraphic traps of the Yanchang Formation has made the Triassic the most important and prolific petroleum play in the Ordos Basin (Jia and Chi, 2004; Zou et al., 2009; Ji et al., 2016). Many scholars have selected Yanchang organic-rich lacustrine mudstones (Chang 7 Shale) as the effective source rocks for the Triassic petroleum play (Zhai, 1997; Hanson et al., 2007; Ji

et al., 2007; Duan et al., 2008; Yu et al., 2010). Two sets of reservoir rocks are present in the Mesozoic petroleum play: the deltaic sandstones of the Upper Triassic Yanchang Formation (Chang 6) and the fluvial sandstones of the Lower Jurassic Fuxian and Yan'an Formation. Although the Triassic petroleum play was discovered almost three decades ago, it continues to attract more and more attention from academic and industrial communities, most recently due to the increasing development of unconventional oil accumulations in ultra-tight reservoirs (Li et al., 2015; Yang et al., 2015; Zou et al., 2015). Some researchers have proposed that crude oils from the Yanchang petroleum play are derived from mixed terrigenous and algal-bacterial organic matter which formed under a reducing and freshwater environment. These oils are predominated by long-chain n-alkanes, abundant C_{29} steranes related to C_{27} and C_{28} homologues, as well as tetracyclic and bicyclic terpanes (Hanson et al., 2007; Duan et al., 2008; Tao et al., 2015). The Chang 7 Shale has been assigned as a principal source rock for generating the Mesozoic oils in the Ordos Basin as it contains abundant bicyclic alkanes that are traced to the most significant biological precursor *B. braunii* (Ji et al., 2007; Ji et al., 2016). In order to better understand the sedimentation of the source rock, and the stack pattern of stratigraphic combination, Liu and Yang (2000) deciphered the sequence stratigraphy of the Upper Triassic - Jurassic sedimentary package based on three unconformity-bounded basin phases. According to Liu et al. (2008), Watson et al. (1987) and Zhao et al. (1996), the Ordos Basin can be subdivided into at least four obvious stages of tectonic deformations existing during the basin's evolution, and the main stages of generation, mineralization and positioning of the multiple energy resources have an obvious connection and relationships with the evolution and reformation of the basin during Mesozoic to Cenozoic period.

Although some scholars have pointed to two Mesozoic intervals as source rocks for the Yanchang petroleum play from nine potential source rock candidates within Proterozoic to Lower Paleozoic marine carbonate, Carboniferous and Permian coal deposits, and Mesozoic lacustrine strata, several important issues regarding aspects of the Triassic petroleum play still need to be further scrutinized based on extensive data and materials. As an expansion of the work presented by Pan et al. (2016), the current study was established to apply depositional environment typing based on geochemical characterization. Here we present a geochemical characterization using a set of samples from the Upper Triassic Yanchang Formation as a case study, in order to take advantage of traditional geochemical methods and the statistical analysis. The simultaneous R- and Q-mode factor analysis as a multivariate statistical approach has been proved effective and useful in grouping samples and discriminating different depositional facies based on tricyclic and pentacyclic terpane compositions (Klovan, 1975; Zhou et al., 1983; Zumberge, 1987; Walden et al., 1992). It allows the variables and

82 samples to be displayed on the same set of diagrams, which greatly facilitates the interpretation of the factors. Alongside
83 the steranes, which contain much useful information on the origin of crude oils (Mackenzie, 1984; Moldowan et al., 1985),
84 terpane biomarkers were chosen as the main statistic variables in the study because of their ubiquitous occurrence, and
85 ease of quantification. Notably, the terpanes tend to be more stable (especially the tricyclics) than steranes when they are
86 subjected to thermal alteration (Seifert and Michael Moldowan, 1978, 1979; Mackenzie, 1984). The detailed summary of
87 the origins of tricyclic and pentacyclic terpanes are outlined by Zumberge (1987). RQ-mode factor analysis is able to
88 separate the investigated samples into different categories according to their different terpane associations. Small
89 maturity fluctuation can influence how the sample behaves, yet the trends do not obscure the interpretation of the
90 genetic sample populations.

91

92 ## 2. Geological framework and thermal history

93 The Ordos Basin is an important non-marine petroliferous basin, and contains oil and natural gas resources that amount
94 to nearly one third of the national annual gross output (Yang and Deng, 2013). As an intraplate depression on the North
95 China Craton during the Mesozoic to Cenozoic, the basin is bounded by a series of synchronous, polyphase orogens, and
96 can be subdivided into six regional structural units (Fig. 1). The stratigraphic combination of the Yanchang Formation
97 consists of ten members named Chang 10 to Chang 1 based on marker beds, sedimentary cycles and lithological
98 association, and is unconformably overlaying and overlain by the Middle Triassic Zhifang Formation and the Lower Jurassic
99 Fuxian Formation, respectively (Fig. 2).

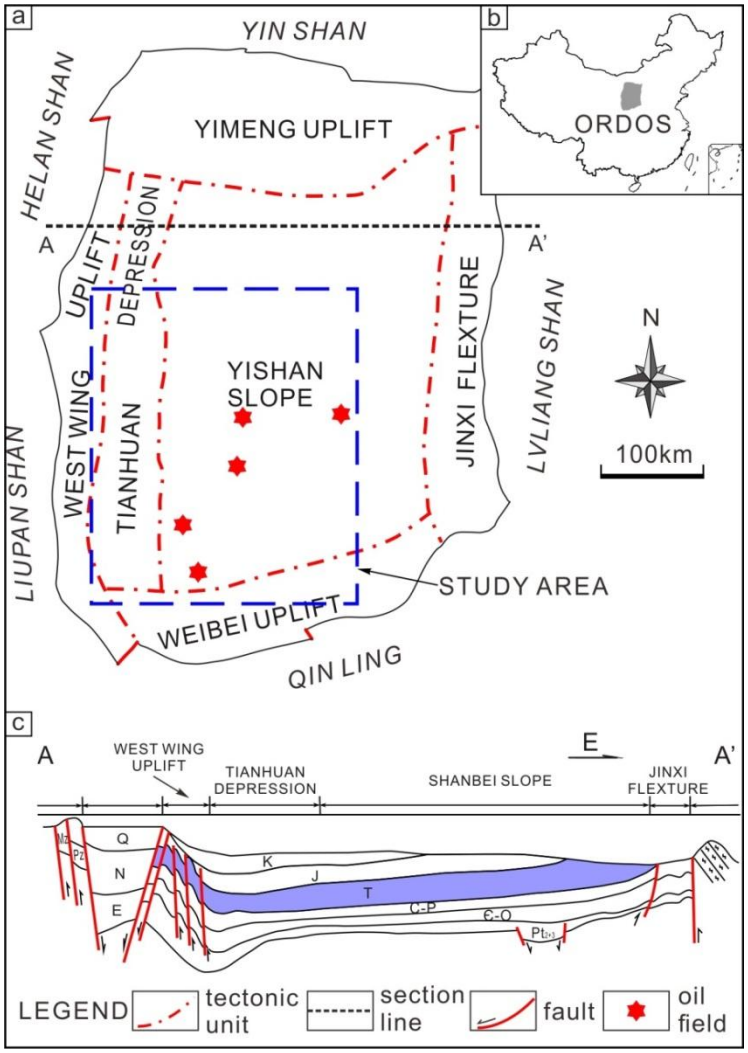


Fig. 1. Sketch map showing tectonic units of Ordos Basin, study area and oil fields locations (a). Inserted map showing the location of Ordos Basin (b). Cross-section across the Ordos Basin illustrating relatively uniform distribution of Triassic stratigraphic packages (c). "Shan" and "Ling" in Chinese both mean mountains.

A detailed analysis of the geological framework and petroleum play of the area under study has recently been published (Pan et al., 2016), building upon earlier works (Sun et al., 1989; Zhao et al., 1996; Liu and Yang, 2000; Yang et al., 2005; Ren et al., 2007; Yang and Deng, 2013; Zou et al., 2013). Tectonic events and thermal history are responsible for the formation of many resources (e.g., petroleum, coal and uranium mine) in the Ordos Basin (Jiao et al., 2005; Chen et al., 2007; Hanson et al., 2007; Ren et al., 2007; Yang et al., 2016). Generally, basin formation since the Paleozoic can be split into three stages (Zhao et al., 1996). (1) From the Early Paleozoic, the basin belongs to the passive continental margin and shelf of the ancient Qinling-Qilian Sea, and is covered by marine Cambrian and Ordovician carbonate rocks. (2) During the

Late Paleozoic through the Early Mesozoic, the basin gradually changes from a marine environment to an inland basin yet a significant marine transgression occurred in the Middle Carboniferous. From the Middle Carboniferous to the Early Permian, coal-bearing sediments have been formed in a littoral setting. From then until the Middle Triassic, non-marine sedimentation characterized by fluvial and lacustrine sandstones and mudstones prevailed throughout the basin. (3) During the Mesozoic, a persistent regression to form an inland basin took place in the Indonesian Movement. Upper Triassic-Lower Cretaceous sedimentary packages are lacustrine organic-rich fine-grained sediments and fluvial clastic rocks bearing several important oil and coal accumulations. In the Early Cretaceous, sedimentation in the basin was halted by the Yanshan Movement. After that (from the Early Cretaceous), the Yanshan and Himalayan Movements led to the deformation of the Paleozoic and Mesozoic stratigraphic packages, and caused basin subsidence to form a broad and asymmetrical north-south-trending syncline with its synclinal axis lying on the west side of the basin.

The Chang 7 Member consists of a set of lacustrine black shales constituting a regional source rock package with a maximum thickness up to 70 m (Yang et al., 2010). Our previous work has documented detailed geochemical properties and predicted organofacies for this shale system (Pan et al., 2016). The Chang 7 Shale consists primarily of laminated shale and occasional tuffaceous mudstone and muddy siltstone interbeds. The strong lithological homogeneity suggests a lack of distinctive cyclicity and a continuous, stable depositional environment during the sedimentation of the Chang 7 Shale. The higher TOC contents (up to ~20%) and the HI values (up to ~450 mg HC/g TOC) assign it as a prolific source rock, and the pseudo van Krevelen diagram indicates a combination of Type I and Type II organic matter. Furthermore, the molecular structures revealed by open system pyrolysis gas chromatography has given detailed insights into the macromolecular composition of the organic matter, and the chain length distribution which can predict the physical properties of the crude oils generated by this particular source rock (Horsfield, 1989). Most of the kerogens from the Chang 7 Shale are enriched in long and moderate alkyl chains (> C_{15+}) and have a high wax oil generation-trend in nature. Other kerogens tend to be dominated by a short to moderate chain length distribution (usually less than C_{14}), ostensibly related to a terrestrial origin. These shale samples are assigned into gas and condensate organofacies, and are likely to generate lighter petroleums with abundant C_1-C_5 hydrocarbons. The basinwide samples used in that study have predicted a predominately Paraffinic High Wax oil generation in nature with a minor combination of a generative nature of Paraffinic-Naphthenic-Aromatic (PNA) Low Wax oil. Some samples also inferred a gas and condensate generative trend for the Chang 7 Shale. These light molecular weight products presumably stem from the cracking of short chain length alkylated aromatic moieties.

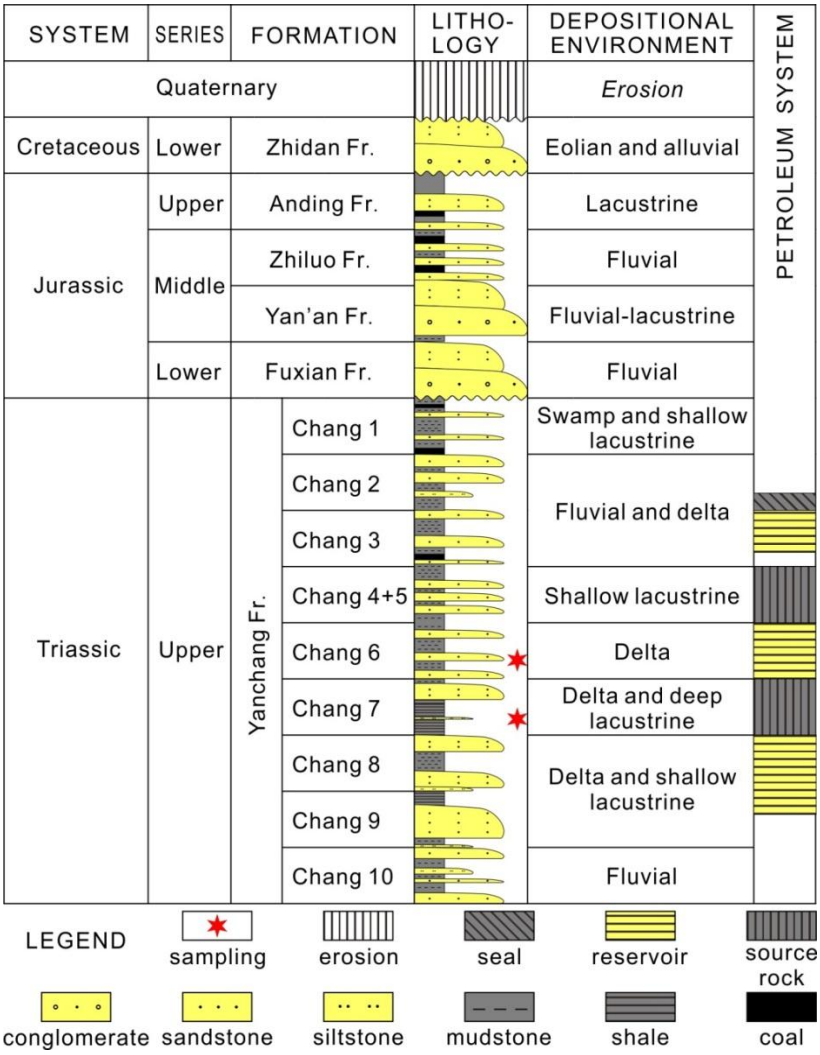


Fig. 2. Stratigraphic columns, depositional environment and source-reservoir-seal associations in the Ordos Basin from Upper Triassic to Quaternary. Modified from (Zhai, 1997; Zou et al., 2012; Zou et al., 2013). The Yanchang Formation has been subdivided into ten members.

Given the deformation and subsidence of the basin from the Early Cretaceous, meaningful thermal activities result in coalification for Permian-Carboniferous coal measures and maturation for the Triassic dispersed organic matter. Before that time, a thermal event occurred around 170-160 Ma in the late Middle Jurassic as a result of subsurface magmatic intrusion in the basin and the magmatic activity in adjacent regions is of significance for intense coalification and hydrocarbon generation. The peak paleotemperatures of the Permian-Carboniferous coal measures amount to greater than 150 ^{o}C, while the Triassic source rocks are reaching 90-160 ^{o}C. After the Late Cretaceous, uplift and erosion activities give rise to the cooling process (Ren et al., 2007). Although others argue for a cooling event began at least approximately since the Miocene due to uplift and erosion in response to the rise of the Qinhai-Tibet Plateau related to the India-Asia

collision in the Himalayan orogeny (Zhao et al., 1996), there is no obvious subsidence of the basin or no significant thermal event that could further cook the petroliferous Triassic system.

3. Materials and methods

3.1. Sample set

The sample set consists of 30 crude oils, 25 potential source rocks (shales) and 13 reservoir sandstones from an Upper Triassic lacustrine-fluvial stratigraphic package, in the Ordos Basin, Central China. They are selected from the Yanchang Formation based on the known depositional environment and petroleum play combinations (Table 1). Specifically, the source rock samples are collected from the Chang 7 Member whereas the reservoir sandstones are sampled from the Chang 6 Member (Fig. 2). Crude oils obtained from exploitation practices are likely to be mixtures of generated hydrocarbons from the various source litho- and organofacies.

3.2. Extraction and fractionation

Shale and reservoir sandstone were powdered (< 0.18 mm) under ambient conditions (<25°C). The powdered materials (roughly 50-150 g for shale and 200 g for sandstone) were subjected to Soxhlet extraction for 24 h at 85 °C using chloroform as solvent in 99.8 % purity (Soxhlet, 1879). After extraction, approximately 20-50 mg extraction yields were weighed in the thermally cleaned vials and separated into maltenes (*n*-hexane-soluble fraction) and asphaltenes by asphaltene precipitation. The condensed extracts were transferred to a pre-cleaned C_{18} non-end capped solid-phase extraction cartridge (Zhang and Huang, 2005). The light fraction which transferred directly through the column along with the residual hydrocarbon from the column eluted with *n*-hexane were collected in the thermally cleaned vial and weighed. Afterward, the polar fraction was eluted with dichloromethane till the droplets were clear. The maltene fraction was further separated into saturated hydrocarbons, aromatic hydrocarbons in a silver nitrate-silica gel solid-phase extraction column using *n*-hexane and dichrolomethane, respectively (Bennett and Larter, 2000). Moreover, roughly 20-50 mg crude oils were directly subjected to asphaltene precipitation and fractionation according to the aforementioned procedures.

3.3. Gas chromatography (GC)

Gas chromatography was performed on saturated hydrocarbon fractions of crude oils and extracts using a Hewlett-Packard 6890N instrument (Agilent Technologies, Waldbronn, Germany), equipped with a flame ionization detector (FID)

and a methyl silicone phase HP DB-5 capillary column with 30 m long, 0.25 mm inner diameter and 0.25 µm film thickness (Mode J&W 122-5-32). The injection device was a Programmed-Temperature Vaporizing (PTV) injector (Agilent Technologies/Gerstel, Mülheim an der Ruhr, Germany) cooled with CO_2 in the temperature of 250 °C. Helium was used as the carrier gas at a constant flow rate of 1.0 mL/min. Samples were injected in cooled on-column mode, and the oven temperature was programmed from 40 to 310 °C at 6 °C /min with initial and final holding times of 2 min and 40 min, respectively. Peak areas of all identifiable *n*-alkanes and acyclic isoprenoids pristane and phytane have been determined using a GC data system.

3.4. Gas chromatography - mass spectrometry (GC-MS)

The saturate fractions of crude oils, shale extracts and reservoir sandstone extracts were analysed by gas chromatography - mass spectrometry using an Agilent 6890N gas chromatograph interfaced to an Agilent 5973 mass selective detector (MSD) utilising electron impact mode (70 eV) with a source temperature of 200 °C. Full-scan mass spectra were recorded from m/z 50 to 500 Da. The capillary column was an Equity-5 (60 m long, 0.25 mm inner diameter and 0.25 µm film thickness), and the carrier gas (helium) was maintained at a constant flow rate of 1.0 mL/min. The GC oven was programmed from 50 to 200 °C at 4 °C /min with an initial holding time of 5 min, then from 200 to 320 °C at 2 °C /min with a final holding time of 25 min. Both interface and injector temperatures were 300 °C, and the transfer line temperature was 250 °C.

No internal standard was added to the samples prior to fractionation, so that no absolute concentration of each component could be achieved during calculation. The quantification of individual compounds was achieved by extracted ion chromatograms and the peak area integration using m/z = 191 to calculate the relative proportions of each terpenoid in Table 1 as well as using m/z = 217 to identify steroids. The compounds were identified based on relative retention times and comparison with published data (Zhang and Huang, 2005; Zhang et al., 2011; Huang et al., 2015).

3.5. Reference dataset

The worldwide data of Zumberge (1987), consisting of 216 crude oils from 27 countries. These oils were grouped into five categories based on the known or suspected depositional facies or other characteristics of the source rocks: 1) deeper-water marine facies; 2) paralic/deltaic nearshore facies; 3) lacustrine facies; 4) phosphate-rich source rocks; and 5)

202 Ordovician-age source rocks. A combined data set of these reference data and those from the Ordos Basin was used in
203 factor analysis, and further in depositional facies classification.

Table 1. Sample types and relative percentage of terpane distributions for crude oils, shale extracts and reservoir sandstone extracts from Yanchang Formation. Q-mode relative contributions to sample loadings are given in the first two factor axes factor-1 and factor-2. The data for groups 1, 2, 3, 4 and 5 crude oils are not listed here, instead readers can access these data from Zumberge (1987).

Type	Sample ID	Tricylic terpanes (%)								Pentacyclic terpanes (%)								Factor scores	
		C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}	C_{31}	C_{32}	OL	GA	Factor-1	Factor-2
oil	R104731	3.66	7.34	12.44	3.56	21.00	16.97	13.90	21.14	26.42	5.54	18.45	32.03	9.70	5.20	0.00	2.66	-0.49	1.38
oil	R104732	5.97	10.13	16.13	3.47	20.65	16.86	11.27	15.51	27.64	7.99	14.37	30.79	10.83	6.05	0.00	2.33	-0.81	1.09
oil	R104733	6.17	10.16	15.66	2.91	20.62	16.26	11.38	16.84	19.36	4.26	20.68	38.13	10.57	5.17	0.00	1.82	-0.76	0.96
oil	R104734	5.32	9.45	15.28	3.49	20.23	16.43	11.67	18.12	20.17	5.43	16.66	38.09	11.76	5.81	0.00	2.09	-0.72	1.08
oil	R104735	5.07	10.37	16.08	3.76	20.76	13.16	9.70	21.11	13.43	2.35	20.71	42.25	13.10	6.69	0.00	1.48	-0.77	0.59
oil	R104736	5.64	11.46	15.67	3.79	19.44	14.39	10.11	19.49	15.55	2.57	22.54	40.81	11.68	5.64	0.00	1.20	-0.84	0.59
oil	R104737	6.69	10.49	15.51	3.53	20.60	16.53	10.85	15.80	28.84	9.19	13.46	30.52	10.21	5.46	0.00	2.32	-0.94	1.06
oil	R104738	5.77	9.79	14.69	2.92	19.94	14.63	12.39	19.87	15.42	2.87	21.65	42.49	11.18	5.34	0.00	1.05	-0.75	0.91
oil	R104739	6.16	10.82	15.82	3.18	19.46	15.48	11.15	17.92	18.95	4.51	19.41	38.36	10.94	5.63	0.00	2.21	-0.85	0.88
oil	R104740	12.10	14.51	19.03	4.09	18.61	13.65	8.90	9.11	62.84	17.82	7.07	12.28	0.00	0.00	0.00	0.00	-2.40	0.86
oil	R104741	3.58	10.00	17.48	2.68	20.37	16.11	11.20	18.59	17.26	2.68	24.60	39.44	9.91	3.84	0.00	2.27	-0.86	1.11
oil	R104742	4.86	12.20	16.36	2.91	19.51	11.98	9.69	22.50	13.07	1.49	25.94	41.40	11.67	5.44	0.00	0.99	-1.01	0.47
oil	R104743	4.39	8.50	12.42	5.06	19.77	17.01	14.99	17.86	19.67	5.57	11.78	38.06	14.10	9.45	0.00	1.37	-0.13	1.05
oil	R104744	6.22	10.11	16.45	3.61	20.28	16.40	11.66	15.28	41.57	11.90	10.70	20.17	7.67	5.16	0.00	2.82	-1.14	1.26
oil	R104745	3.57	9.63	17.12	3.95	20.95	15.22	10.97	18.60	19.64	4.43	24.17	35.96	9.48	4.65	0.00	1.67	-0.74	0.94
oil	R104746	6.98	10.05	16.40	4.12	20.53	15.21	10.47	16.24	27.81	7.24	16.58	30.30	10.44	5.68	0.00	1.95	-0.88	0.81
oil	R104747	4.80	9.81	17.55	3.08	20.57	15.16	10.43	18.60	16.66	3.32	22.17	40.91	10.68	4.71	0.00	1.55	-0.89	0.96
oil	R104748	4.94	10.56	17.26	3.10	19.86	15.24	10.62	18.41	17.18	3.07	21.96	40.59	10.80	4.47	0.00	1.94	-0.93	0.91
oil	R104749	4.97	10.10	16.85	3.41	20.32	15.06	11.03	18.27	18.05	3.34	20.66	40.08	10.23	5.16	0.00	2.48	-0.86	0.93
oil	R104750	4.79	10.41	17.05	3.50	20.06	15.24	10.81	18.15	17.49	3.60	21.63	40.07	10.18	4.79	0.00	2.25	-0.87	0.91
oil	R104751	5.02	10.51	16.55	3.04	20.90	16.13	11.03	16.82	18.51	4.42	20.69	38.97	10.69	5.14	0.00	1.57	-0.80	0.98
oil	R104752	3.30	11.58	19.44	2.68	20.87	13.11	9.61	19.40	13.34	1.13	28.10	43.06	9.03	3.54	0.00	1.80	-1.10	0.76
oil	R104753	7.28	10.84	16.98	4.13	19.86	16.43	10.56	13.91	45.39	14.81	14.02	25.78	0.00	0.00	0.00	0.00	-1.78	1.46
oil	R104754	5.48	10.35	16.72	3.93	20.86	16.84	11.28	14.54	39.67	11.39	18.24	30.70	0.00	0.00	0.00	0.00	-1.55	1.50
oil	R104755	5.19	10.61	16.97	3.19	20.10	14.13	10.14	19.67	14.16	2.17	23.56	43.41	10.78	4.65	0.00	1.27	-0.94	0.77
oil	R104756	5.71	10.99	17.67	3.05	20.40	16.40	11.04	14.75	34.72	10.02	18.52	26.49	7.72	0.00	0.00	2.53	-1.33	1.27
oil	R104757	4.49	10.10	16.71	3.07	19.93	14.56	10.86	20.28	14.01	2.31	22.78	44.01	10.63	4.51	0.00	1.76	-0.91	0.94
oil	R104758	4.31	10.16	17.36	2.94	19.57	14.71	10.80	20.15	13.49	1.88	23.57	45.04	10.64	4.13	0.00	1.25	-0.97	0.97
oil	R104759	5.45	10.59	16.41	4.12	20.49	16.47	11.38	15.09	30.42	8.75	15.07	27.73	9.65	5.12	0.00	3.26	-0.87	1.06
oil	R104760	5.01	10.27	16.37	3.21	20.51	16.17	10.78	17.68	17.71	3.72	21.16	39.57	10.57	5.46	0.00	1.81	-0.78	0.96
SRE	R104761	6.91	11.53	17.99	3.48	19.96	12.08	9.74	18.30	22.34	4.12	4.07	41.26	15.42	8.82	0.00	3.96	-1.19	0.67
SRE	R104762	6.92	15.33	17.61	4.83	19.32	12.33	10.55	13.10	26.63	7.48	7.63	31.88	13.06	7.39	0.00	5.93	-1.12	0.33
SRE	R104763	10.97	17.67	14.17	6.69	14.23	10.74	9.03	16.51	42.69	4.87	5.71	23.54	11.27	7.71	0.00	4.21	-1.56	-0.39

SRE	R104764	8.92	15.75	15.90	4.15	17.63	12.23	8.77	16.64	34.33	4.66	5.53	33.22	11.87	7.01	0.00	3.38	-1.59	0.25
SRE	R104765	9.14	13.16	17.82	3.20	16.59	11.11	7.87	21.11	23.81	2.12	2.95	40.58	16.46	12.13	0.00	1.95	-1.43	0.24
SRE	R104766	7.59	16.81	19.79	4.10	17.74	11.01	7.79	15.17	30.45	4.78	6.83	34.93	13.57	6.83	0.00	2.62	-1.72	0.13
SRE	R104767	7.73	14.21	19.80	2.96	19.08	11.54	7.51	17.18	32.25	4.54	5.91	38.68	10.57	5.75	0.00	2.31	-1.93	0.59
SRE	R104768	8.05	12.98	17.92	4.25	18.63	13.53	9.86	14.77	28.61	6.00	6.94	33.01	13.84	9.08	0.00	2.53	-1.12	0.45
SRE	R104769	6.67	16.97	21.89	4.31	17.35	10.86	6.95	15.00	22.62	2.65	3.53	41.85	18.70	8.88	0.00	1.77	-1.60	0.00
SRE	R104770	8.32	18.11	17.97	4.30	16.56	11.87	7.91	14.97	31.41	3.84	5.18	35.67	14.88	6.41	0.00	2.61	-1.72	0.03
SRE	R104771	8.03	13.28	15.82	5.19	16.04	12.96	9.97	18.72	27.62	5.55	7.05	31.48	13.61	10.74	0.00	3.95	-1.04	0.29
SRE	R104772	8.27	12.86	16.97	3.09	18.49	13.92	9.53	16.86	35.54	6.16	7.02	31.16	10.33	6.93	0.00	2.86	-1.51	0.75
SRE	R104773	7.69	12.56	18.46	4.32	19.90	15.69	8.57	12.81	40.84	10.89	10.26	19.78	8.21	7.77	0.00	2.26	-1.25	0.71
SRE	R104774	7.78	11.24	17.46	2.94	18.06	14.48	10.16	17.88	31.70	5.89	6.18	35.80	11.44	7.20	0.00	1.79	-1.43	0.96
SRE	R104775	15.63	15.99	16.00	4.84	16.82	13.00	8.59	9.13	26.59	14.86	18.00	20.90	8.34	8.53	0.00	2.77	-1.06	-0.27
SRE	R104776	11.08	14.88	20.70	4.58	17.56	12.80	7.25	11.16	36.27	13.31	14.32	15.23	8.88	8.29	0.00	3.70	-1.39	0.09
SRE	R104777	8.76	14.64	22.90	5.24	18.77	13.39	6.74	9.55	36.04	14.98	14.43	18.23	7.19	6.24	0.00	2.91	-1.52	0.34
SRE	R104778	8.01	13.09	21.64	4.93	19.41	14.72	7.67	10.51	32.40	17.70	17.63	17.42	6.51	4.91	0.00	3.43	-1.31	0.67
SRE	R104779	5.83	9.56	16.17	4.24	20.90	18.19	11.31	13.82	33.39	17.15	14.47	16.61	9.02	6.76	0.00	2.59	-0.64	1.27
SRE	R104780	10.91	15.53	17.53	4.21	19.18	13.14	9.33	10.17	25.36	16.63	13.91	23.07	11.30	7.30	0.00	2.43	-1.07	0.21
SRE	R104781	10.72	15.95	21.08	3.68	18.60	11.36	7.05	11.56	40.17	12.18	14.99	20.80	6.72	3.65	0.00	1.49	-1.94	0.22
SRE	R101399	13.49	15.37	18.24	3.78	16.68	11.16	7.94	13.33	38.27	5.96	6.92	29.32	10.95	5.22	0.00	3.37	-1.88	0.04
SRE	R101400	8.58	13.68	15.50	4.28	16.51	9.81	7.93	23.70	21.37	1.91	2.03	44.25	17.30	10.19	0.00	2.96	-1.46	0.13
SRE	R101408	10.89	13.34	12.65	5.09	14.41	8.37	7.28	27.96	25.49	1.75	1.92	36.72	19.20	12.01	0.00	2.91	-1.36	-0.29
SRE	R101409	17.85	14.33	13.59	3.56	12.48	5.95	5.43	26.81	20.98	2.17	3.23	40.46	21.57	9.94	0.00	1.66	-1.81	-0.74
RE	R101401	7.38	10.90	18.31	3.56	18.79	11.60	8.81	20.64	26.26	2.64	2.81	41.34	15.76	8.35	0.00	2.84	-1.43	0.65
RE	R101402	5.51	11.83	16.43	2.80	19.07	13.06	9.11	22.18	19.02	2.41	2.97	53.29	14.08	6.02	0.00	2.21	-1.59	1.04
RE	R101403	5.01	9.83	14.75	3.26	19.71	14.41	10.75	22.28	18.62	2.41	3.10	48.55	16.40	9.16	0.00	1.76	-0.99	1.05
RE	R101404	4.90	9.64	14.47	3.53	20.17	14.77	11.00	21.51	17.23	2.31	2.77	49.36	16.65	9.50	0.00	2.18	-0.86	1.05
RE	R101405	4.88	9.94	14.66	3.60	20.11	13.76	11.21	21.83	17.74	2.25	2.94	49.73	16.21	9.30	0.00	1.82	-0.94	0.99
RE	R101406	4.24	10.21	15.27	3.44	20.47	13.41	10.80	22.16	16.74	2.01	2.57	51.47	16.61	8.70	0.00	1.89	-1.03	1.02
RE	R101407	4.47	10.90	15.87	2.96	19.95	12.55	9.52	23.77	15.93	1.63	2.14	54.71	16.03	7.80	0.00	1.75	-1.35	1.02
RE	R101410	4.58	11.50	15.99	2.59	20.00	12.44	9.61	23.29	17.27	1.98	2.42	54.99	14.78	6.76	0.00	1.81	-1.50	1.08
RE	R101411	4.68	11.68	16.31	2.94	19.57	12.05	9.25	23.52	16.76	1.73	2.33	54.73	15.36	7.64	0.00	1.45	-1.48	0.95
RE	R101412	4.29	11.16	15.79	2.70	19.71	12.54	9.57	24.24	16.53	1.70	2.21	55.10	15.51	7.13	0.00	1.82	-1.45	1.08
RE	R101413	4.79	11.74	16.16	2.76	19.23	12.35	9.69	23.28	16.79	1.83	2.33	54.64	15.30	7.40	0.00	1.70	-1.48	1.00
RE	R101414	4.32	10.62	15.75	2.90	20.43	12.63	10.07	23.29	16.24	1.77	2.27	54.19	15.95	7.93	0.00	1.63	-1.28	1.05
RE	R101415	4.85	10.82	16.18	3.15	19.23	14.11	9.87	21.79	19.24	2.47	2.94	50.21	15.29	7.49	0.00	2.35	-1.29	1.09

Note:

Terpane identifications are shown in Fig. 3.

The C_{27} is a combination of the Ts and Tm trisnorhopanes, and C_{31} and C_{32} extended hopanes are combinations from their 22S and 22R stereoisomers.

SRE = Source Rock Extracts; RE = Reservoir Extracts.

3.6. Factor analysis

Simultaneous R- and Q-mode (RQ-mode) factor analysis was applied as a tool for assisting interpretation of terpane biomarkers and reducing the dimensionality (Zhou et al., 1983; Zumberge, 1987; Walden et al., 1992). Either R-mode or Q-mode factor analysis requires a standardization of the raw data matrix by subtracting the mean from each corresponding terpane value and being divided by the standard deviation, to remove the effects of different variables being measured on different scales. As a multivariate statistical technique, the RQ-mode factor analysis extracts a common set of independent factors (i.e., factor-1 and factor-2) that are linear combination of the original variables (i.e., relative terpane abundance). In this scenario, both the variable factor loadings (R-mode) and the sample factor loadings (Q-mode) are relative to those same independently underlying factors. Thus, the variables and samples are scaled so they can be plotted on the same set of factor axes to uncover the interrelationships among the variables and to find grouping distributions of samples.

The analyses in this study were performed using IBM SPSS Statistics software (Ver. 22). Principal components analysis rather than correspondence analysis was employed as an extraction method because the correlation coefficient or covariance can better measure the similarities between variables while the Euclidean distance tends to represent the similarities between samples (Zhou et al., 1983). The RQ-mode analysis avoids involving any rotation procedure, as criticized by Temple (1978), in applying factor analysis in geology. Both R- and Q-mode factor analyses are based on eigenvector methods and can be performed separately (Walden et al., 1992). Samples with similar terpane distributions were grouped together in two-dimensional space defined by a set of factor axes, and specific variables which are of importance in distinguishing the groups were determined.

4. Results

4.1. R-mode factor analysis

Eight tricyclic (C_{19} - C_{26}) and eight pentacyclic (C_{27} - C_{32}, oleanane and gammacerane) terpanes have been identified in the m/z = 191 mass chromatogram from saturated fractions in Zumberge (1987)'s database and in this study (Fig. 3). Oleanane is absent in the Yanchang Formation because it is highly specific for higher-plant input of Cretaceous or younger age. The eight tricyclic terpanes are single components for the most part, but some show coelution between their isomers or other components. For example, the 22S and 22R stereoisomers of C_{25} triterpane coelute at 51.5 retention time, and the

tetracyclic terpane elutes immediately prior to the two C_{26} stereoisomers at about 54 min. In contrast to the tricyclics, some of the pentacyclic components are combined and used as a single variable. For example, the C_{27} is a combination of the Ts and Tm trisnorhopanes to minimize the effect of thermal maturation (Zumberge, 1987), and C_{31} and C_{32} extended hopanes are combinations from their 22S and 22R stereoisomers. All relative percentages of terpane components from the Yanchang Formation database are listed in Table 1.

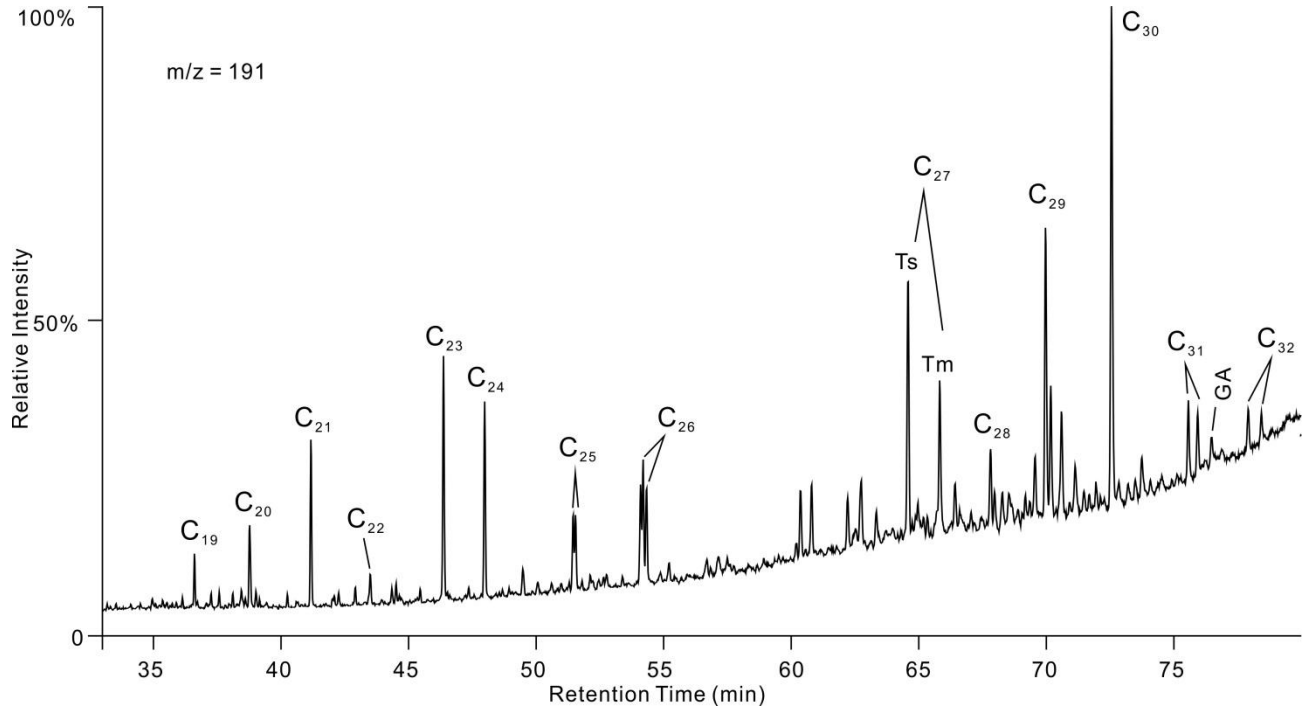


Fig. 3. Mass fragmentogram (m/z = 191) of the saturated fraction from a crude oil sample (R104731) showing the 16 terpane variables used in this study. From C_{19} to C_{26} are tricyclic terpanes. C_{27} is a combination of Ts and Tm, i.e., 18α(H)-22,29,30-trisnorneohopane and 17α(H)-22,29,30-trisnorhopane. C_{28} is 17α(H),21β(H)-29,30-bisnorhopane. C_{29} is 17α(H),21β(H)-30-norhopane. C_{30} is 17α(H),21β(H)-hopane. GA is gammacerane. C_{31} consists of the (22S)- and (22R)-17α(H),21β(H)-29-homohopanes, whereas C_{32} is composed of (22S)- and (22R)-17α(H),21β(H)-29-bishomohopanes.

Mean and standard deviations of the 16 terpane variables from a combination data set of 68 samples in Table 1 and 216 crude oil samples from Zumberge (1987) are shown in the upper part of Table 2, and a single data set of samples from the Yanchang Formation is listed in the lower part. It should be emphasized here that the database of 216 crude oils was utilized during the whole factor analysing procedure in this study, which means this database coupled with the data set from Yanchang Formation both make a contribution to the factor loadings in R-mode factor analysis. Only with this consideration could the sample grouping for the Yanchang Formation be comparable with the sample distributions of

Zumberge (1987)'s results in Q-mode factor analysis. Table 2 and Fig. 4 show that the C_{23} and C_{30} compounds are the largest tricyclic and pentacyclic components, respectively, both for the combined dataset and for the single dataset of the Yanchang Formation. Furthermore, the C_{26} and C_{27} terpanes from a single dataset rank only second to C_{23} and C_{30} compounds, while for a combined dataset, C_{21} and C_{29} take second place in corresponding tricyclic and pentacyclic components (Fig. 4).

Table 2. Mean and standard deviations of the 16 terpane variables.

		C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}	C_{31}	C_{32}	OL	GA
combined data set	Mean (%)	7.28	10.86	14.14	5.07	22.15	13.97	11.60	9.59	14.56	3.12	17.96	27.43	13.37	8.21	1.55	4.29
	Min.	0.30	3.50	5.60	0.00	5.00	2.40	1.00	1.00	4.20	0.00	1.92	11.20	0.00	0.00	0.00	0.00
	Max.	52.00	31.70	25.80	11.80	42.60	20.50	21.30	27.96	62.80	22.80	40.40	55.10	27.80	15.30	36.50	31.20
	Std. Dev.	7.68	4.02	3.52	1.85	6.00	2.91	3.61	5.44	8.46	4.58	6.36	8.52	4.17	2.95	5.23	4.79
Yanchang single set	Mean (%)	6.79	11.93	16.83	3.68	19.19	13.79	9.86	17.94	25.00	5.72	12.08	36.85	11.75	6.41	0.00	2.20
	Min.	3.30	7.30	12.40	2.59	12.48	5.95	5.43	9.10	13.10	1.10	1.92	12.30	0.00	0.00	0.00	0.00
	Max.	17.85	18.11	22.90	6.69	21.00	18.19	15.00	27.96	62.80	17.80	28.10	55.10	21.57	12.13	0.00	5.93
	Std. Dev.	2.85	2.39	2.04	0.79	1.74	2.25	1.66	4.25	9.84	4.63	8.10	10.85	4.08	2.53	0.00	0.97

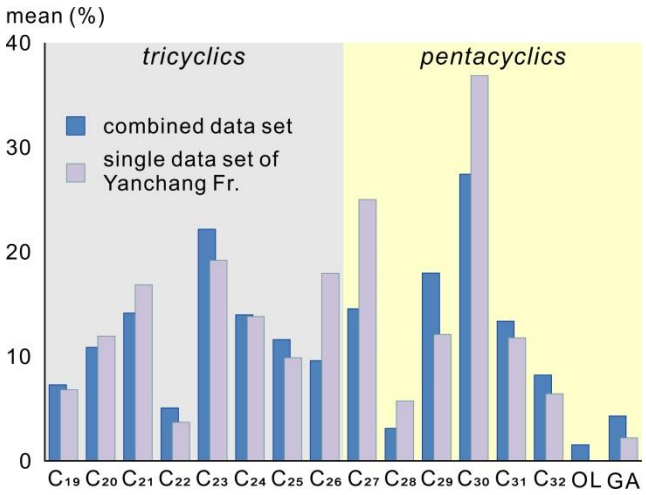


Fig. 4. Mean value distributions for a combined data set and a single data set of Yanchang Formation.

The linear correlation coefficients for each combination of variables are given in the correlation matrix (Table 3). Regarding the tricyclic terpanes, C_{19} and C_{20} compounds have the highest positive correlation coefficient (i.e., 0.62), whereas these two lower molecular weight terpanes negatively correlate to the relatively higher molecular weight compounds from C_{23} to C_{26}. Each terpane usually has a relatively high positive correlation with its adjacent compound from C_{19}, irrespective of C_{21} and C_{22} which show a negative correlation coefficient of - 0.40. C_{23}, C_{24} and C_{25} compounds

exhibit positive correlations with each other, and C_{24} and C_{25} have the highest coefficient (0.66). For the pentacyclic terpanes, higher molecular weight C_{31} and C_{32} are the strongest related compounds (0.75), not only for the pentacyclics but for the whole terpanes. This strong correlation perhaps stems from their common origin (Zumberge, 1987). Norhopane (C_{29}) and hopane (C_{30}) show the second strongest but nevertheless negative correlation (-0.46) in the pentacyclic terpanes. Moreover, it should be noticed that tricyclic C_{26} has a very high correlation coefficient (0.79) with pentacyclic C_{30}, which is almost twice as much as that value when only considering the Zumberge (1987)'s dataset. This reflects that the correlation between these two terpanes in the Yanchang Formation data set is much stronger than that in the dataset of Zumberge (1987). Nevertheless, there is no clear relationship on specific compounds through tricyclic to pentacyclic terpanes, but a generally negative correlation between lower molecular weight tricyclics and higher molecular weight pentacyclics can be seen.

Table 3. Correlation matrix of terpane variables. Absolute linear correlations discussed in the paragraph are highlighted.

	C_{19}	C_{20}	C_{21}	C_{22}	C_{23}	C_{24}	C_{25}	C_{26}	C_{27}	C_{28}	C_{29}	C_{30}	C_{31}	C_{32}	OL	GA
C_{19}	1.00															
C_{20}	**0.62**	1.00														
C_{21}	-0.19	**0.29**	1.00													
C_{22}	0.01	-0.13	**-0.40**	1.00												
C_{23}	**-0.61**	**-0.63**	-0.38	**0.41**	1.00											
C_{24}	**-0.65**	**-0.57**	0.09	-0.21	**0.27**	1.00										
C_{25}	**-0.57**	**-0.67**	-0.29	-0.05	**0.40**	**0.66**	1.00									
C_{26}	-0.20	0.00	0.41	-0.50	-0.26	0.08	**-0.09**	1.00								
C_{27}	0.16	0.25	0.32	-0.16	-0.27	-0.09	-0.38	0.41	1.00							
C_{28}	-0.14	-0.11	0.10	-0.23	0.10	0.24	0.13	0.13	0.32	1.00						
C_{29}	-0.03	-0.11	-0.26	0.34	0.32	0.04	0.05	-0.56	-0.32	-0.28	1.00					
C_{30}	-0.04	0.09	0.32	-0.44	-0.29	-0.06	-0.15	**0.79**	0.09	-0.17	**-0.46**	1.00				
C_{31}	0.01	-0.22	-0.43	0.32	0.27	-0.21	0.12	-0.11	-0.33	-0.38	0.02	0.10	1.00			
C_{32}	0.01	-0.28	-0.40	0.33	0.24	-0.11	0.24	-0.26	-0.36	-0.26	-0.05	-0.15	**0.75**	1.00		
OL	0.20	0.26	-0.07	0.08	-0.10	-0.09	-0.16	-0.19	-0.09	-0.06	0.08	-0.18	-0.35	-0.32	1.00	
GA	-0.20	-0.07	0.23	-0.02	0.01	0.15	0.16	-0.12	-0.28	-0.08	-0.11	-0.19	-0.24	-0.06	-0.10	1.00

Table 4 gives eigenvalues, total variance (%) and cumulative total variance (%) for a combined data set. The first two factors with eigenvalues of 4.11 and 3.13, account for over 45 percent of cumulative total variance in the data set. The remaining factors are described as around 13%, 9% and 7%, respectively, to share the residual total variance. Thus, two factors should be extracted to create a two-dimensional R-mode factor analysis, and the relative contributions of each variable to factor loadings are listed in Table 5. The length of the vector represents the importance of the variable in describing the variance in the data, whereas the length of the vector projection of a variable on the axis reflects the

contribution of this variable to the corresponding factor loading. The more the variable vector parallels to the factor axis, the higher the axis gains relative factor loadings for this particular variable (Fig. 5). For example, the factor-1 is composed of tricyclic C_{20}, C_{21}, C_{23} and C_{25} as well as two pentacyclic C_{27} and C_{32} which account for 52% cumulative factor loadings, while the factor-2 consists of most tricyclic terpanes, such as C_{19}, C_{20}, C_{22}, C_{24}, C_{25} and C_{26}, covering 57% contribution to factor loadings (Table 6). Some variables, such as C_{20} and C_{25}, have approximately equal loadings on both factor-1 and factor-2. In total, the tricyclic terpanes are more important than pentacyclic terpanes for both two factors loadings in describing the variance in the data set. The total contributions of tricyclics to factor-1 and factor-2, for example, are 60.20% and 65.64%, respectively, whereas these percentages for pentacyclics are only 39.80% and 34.36% (Table 5).

Table 4. Eigenvalues, total variance (%) and cumulative total variance (%).

Factor	Eigenvalue	Total variance (%)	Cumulative total variance (%)
1	4.11	25.67	25.67
2	3.13	19.54	45.22
3	2.20	13.77	58.98
4	1.40	8.73	67.72
5	1.11	6.91	74.62

Table 5. R-mode relative contributions to factor loadings. The abbreviation "F" refers to "Factor".

	F-1	%	F-2	%	F-3	F-4	F-5
C_{19}	-0.43		-0.75		0.04	0.07	0.12
C_{20}	-0.69		-0.54		-0.10	-0.13	0.07
C_{21}	-0.61		0.37		-0.10	-0.31	0.12
C_{22}	0.53	60.20	-0.49	65.64	-0.08	0.16	0.09
C_{23}	0.76		0.21		-0.07	0.25	-0.10
C_{24}	0.32		0.76		-0.24	-0.04	-0.11
C_{25}	0.63		0.56		-0.01	-0.07	-0.02
C_{26}	-0.56		0.55		0.46	0.08	-0.19
C_{27}	-0.59		0.12		-0.05	0.57	0.20
C_{28}	-0.15		0.43		-0.32	0.57	0.36
C_{29}	0.46		-0.33		-0.41	-0.05	-0.36
C_{30}	-0.49	39.80	0.31	34.36	0.63	-0.15	-0.38
C_{31}	0.49		-0.29		0.74	0.04	0.01
C_{32}	0.57		-0.25		0.58	-0.01	0.30
OL	-0.15		-0.28		-0.51	-0.04	-0.48
GA	0.09		0.21		-0.26	-0.70	0.50

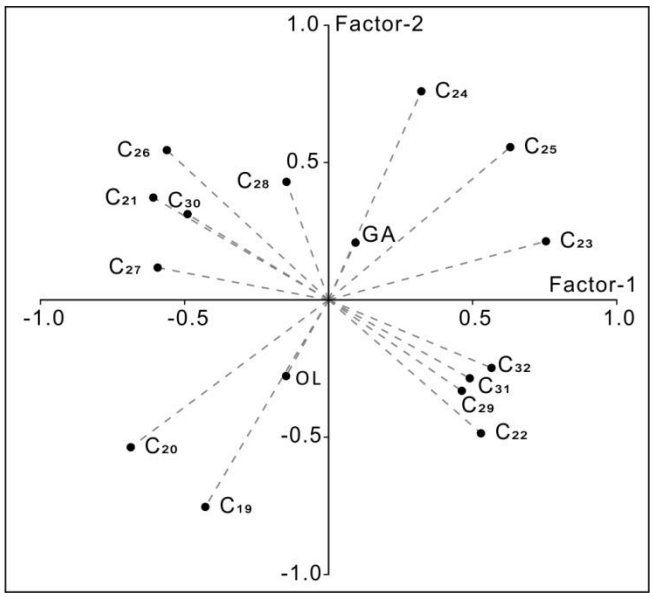


Fig. 5. R-mode factor analysis showing the interrelationship of the 16 terpane variables in factor- 1 and 2 space. The compositions of factors- 1 and 2 are given in Table 5. These two factors account for over 45 percent of the variation in the data.

4.2. Molecular composition

Selected GC-MS-derived biomarker and non-biomarker ratios from hopane compounds and aromatic fractions, respectively, are used to evaluate thermal maturity. Common maturity-related isomerization ratios, i.e. hopane C_{32} 22S/(22S + 22R) and sterane C_{29} 20S/(20S + 20R), have average values of 0.57 and 0.52, respectively. These parameters, however, are not valid for evaluating the thermal maturity of the samples in this study, as they have reached the equilibrium values, i.e., homohopane isomerization equilibrium = 0.57-0.62, and ethyl-cholestane isomerization equilibrium = 0.52-0.55 (Seifert and Moldowan, 1986). Instead, alkyl homologs of the polycyclic aromatic hydrocarbon (PAH) are thought to be largely indicative of thermal evolution of the source rock and oils at the time of expulsion (Radke et al., 1982; Alexander et al., 1984; Radke et al., 1984; Alexander et al., 1985; Radke et al., 1986; Radke, 1988). This is because that the PAHs appear to be anti-biosynthesis (Hase and Hites, 1976) and their complex assemblages and alkyl homologs are at least partly derived from non-aromatic biogenic precursors (Laflamme and Hites, 1978; Wakeham et al., 1980).

Table 6. Absolute contributions of variables to each factor loading, and the cumulative percentage.

Factor-1	loadings (%)	cumulative (%)	Factor-2	loadings (%)	cumulative (%)
C_{23}	10.06	10.06	C_{24}	11.80	11.80
C_{20}	9.14	19.20	C_{19}	11.72	23.52
C_{25}	8.41	27.60	C_{25}	8.64	32.17
C_{21}	8.10	35.70	C_{26}	8.47	40.64
C_{27}	7.90	43.60	C_{20}	8.35	48.99
C_{32}	7.54	51.14	C_{22}	7.56	56.55
C_{26}	7.47	58.61	C_{28}	6.67	63.22
C_{22}	7.05	65.66	C_{21}	5.78	69.00
C_{31}	6.53	72.19	C_{29}	5.16	74.16
C_{30}	6.51	78.70	C_{30}	4.85	79.01
C_{29}	6.17	84.87	C_{31}	4.45	83.46
C_{19}	5.69	90.56	OL	4.32	87.78
C_{24}	4.29	94.84	C_{32}	3.86	91.64
OL	1.96	96.80	C_{23}	3.31	94.95
C_{28}	1.94	98.75	GA	3.23	98.18
GA	1.25	100.00	C_{27}	1.82	100.00

Three Methyldibenzothiophene Ratios, i.e., MDR1, MDR2,3 and MDR4, coupled with a maturity-related biomarker Ts/(Ts + Tm) have been used in mirroring the maturity evolution for crude oils and rock extracts. These sulphur-containing aromatic ratios mirror the aromaticity of the aromatic fraction, which has been shown to increase with increasing depth of burial, i.e., maturity level (Tissot et al., 1971). The methyl homologs are formed through alkylation of the parent compound, i.e., the dibenzothiophenes (Radke et al., 1982). On the other hand, although the Ts/(Ts + Tm) ratio is most reliable as a maturity indicator owing to a nature of Ts being thermally more active than Tm during catagenesis (Seifert and Michael Moldowan, 1978), it should be used with caution when estimating oils sharing a common source of consistent organic facies, because the Ts/(Ts + Tm) appears to be sensitive to clay-catalysed reactions (Peters et al., 2007). Here, we report a combined result of maturity-related parameters avoiding the misevaluation from different types of samples.

Fig. 6 displays the distribution of these parameters according to a sample numbering system in Table 1 rather than to burial depth. Sample groups from the Yanchang Formation, i.e., crude oils, source rock and reservoir extracts, are discriminated by dotted lines. Three ratios, MDR1, MDR2,3 and Ts/(Ts + Tm), possess a similar vertical profile in which they show minor fluctuation. This indicates a similar maturity level for all Yanchang Formation samples. The third methyl dibenzothiophene ratio (MDR4), however, doesn't appear to follow the pattern of consistent maturity level but shows a very high level in source rock extracts and relatively lower maturity in crude oils and reservoir extracts. This will be discussed below. For methyl dibenzothiophene ratios, the absolute values of MDR1, MDR2,3 and MDR4 present a trend in

which MDR1 has the lowest values (mean 0.39) compared to the highest MDR4 values (mean 2.25) with MDR2,3 possessing intermediate values (mean 1.13). This phenomenon is also supported by what has been observed in electrophilic substitution when considering the difference of the reaction behaviour of dibenzothiophene, i.e., the 4- and 2-position of dibenzothiophene are more active than the 1-positon (Hartough and Meisel, 1954; Radke et al., 1982).

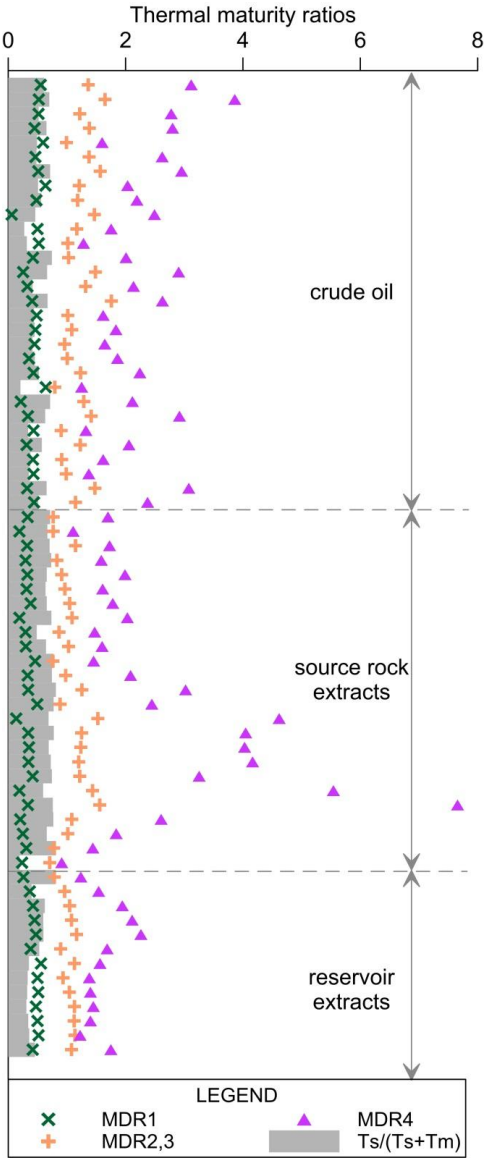
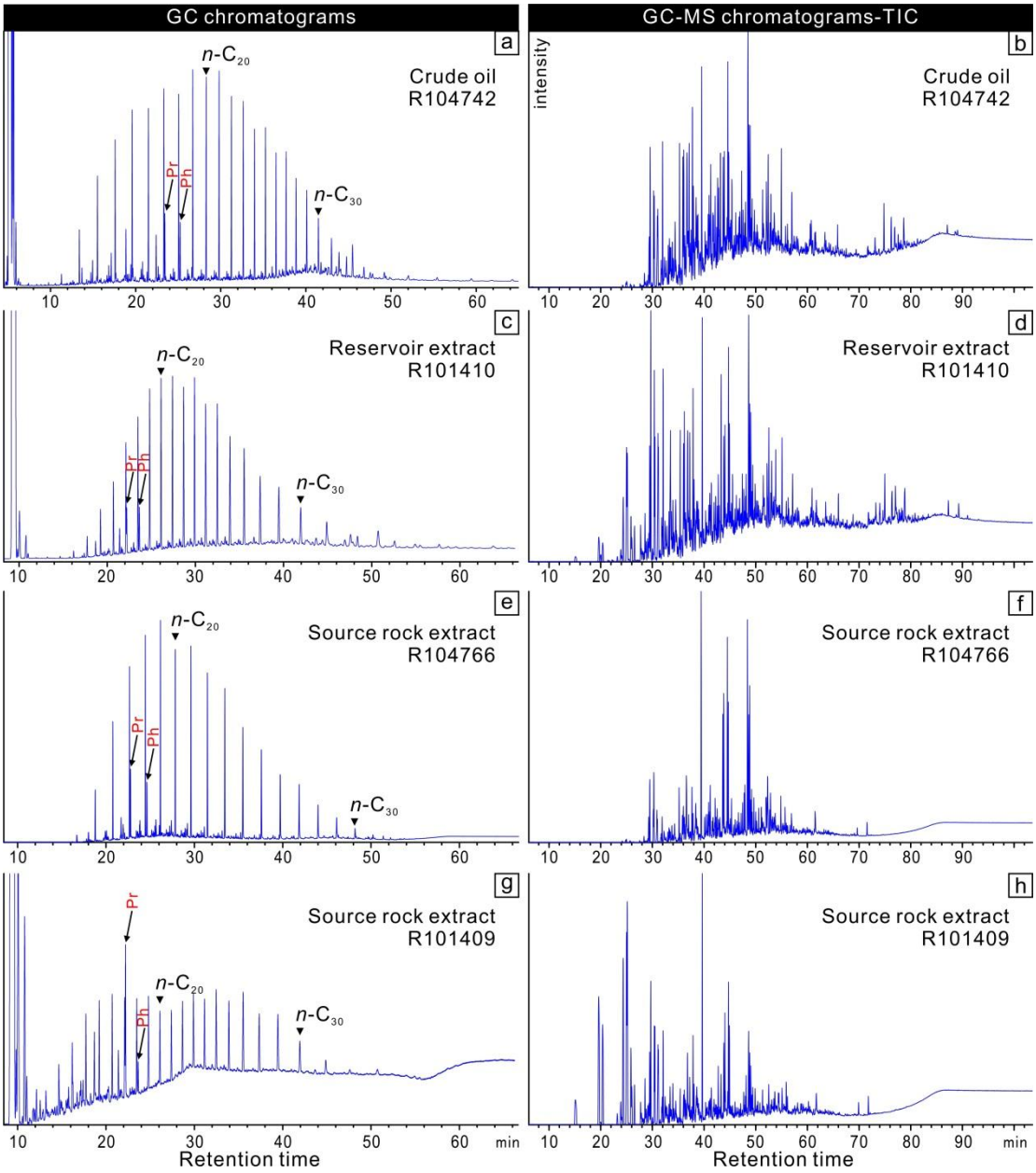


Fig. 6. Distribution of maturity-related parameters derived from C_{27} trisnorhopane homologs (Ts and Tm) and alkyl dibenzothiophene (DBT) homologs. Ts = 18α(H)-22,29,30-trisnorneohopane; Tm = 17α(H)-22,29,30-trisnorhopane. MDR is the abbreviation of Methyl Dibenzothiophene Ratio. MDR1 = 1-MDBT/DBT; MDR2,3 = (2-MDBT + 3-MDBT)/DBT; MDR4 = 4-MDBT/DBT. Arrangement of samples sequence from the top to the bottom isn't related to the depth but to the sample numbering system shown in Table 1.

345 Some GC-derived isoprenoid biomarkers and GC-MS-determined sterol homologs are here presented for revealing the
346 genetic correlation between source-reservoir and reservoir-oil associations. The relationships between the isoprenoids
347 pristane and phytane as well as their adjacent *n*-alkanes have been widely used as molecular indices for assessing
348 depositional environment and organic matter type (Brooks et al., 1969; Powell and McKirdy, 1973; Didyk, 1978; Powell,
349 1988; Peters et al., 1999; Hill et al., 2007). A diagram of pristane/*n*-C_{17} versus phytane/*n*-C_{18}, utilizing the GC data depicted
350 in Figs. 7a, c, e and g, is shown in Fig. 8. Most of the studied samples plot within the region of mixed organic matter
351 deposited under reducing conditions. The organic matter is of Type II or a mixed Type II and III, which is consistent with
352 our previous work upon the characterization of geochemistry and organofacies for Ordos lacustrine source rocks showing
353 an aquatic origin of organic precursors (Pan et al., 2016). Only one sample (R101409) from the source rock extracts shows
354 a composition more typical of a more oxidizing depositional environment with significantly high intensity of pristane (Fig.
355 7g). Interestingly, for each sample group, reservoir extracts appear to have been subjected to more intensive oxidization
356 compared to crude oils and source bitumens, while ratios for crude oils appear to indicate a more consistent anoxic
357 condition. Source bitumens, however, are span a wide range of mixed anoxic, dysoxic depositional conditions irrespective
358 of kerogen type.



359

360 **Fig. 7.** Selected GC-derived chromatograms and GC-MS-derived TIC chromatograms for crude oils, reservoir extracts and source bitumens from saturated

361 and aromatic fractions, respectively. Isoprenoids pristane and phytane as well as carbon numbers 20 and 30 of *n*-alkanes are marked in GC

362 chromatograms.

363

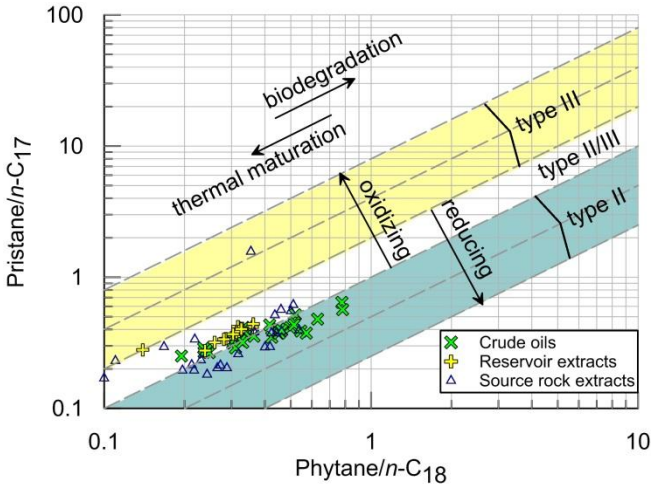
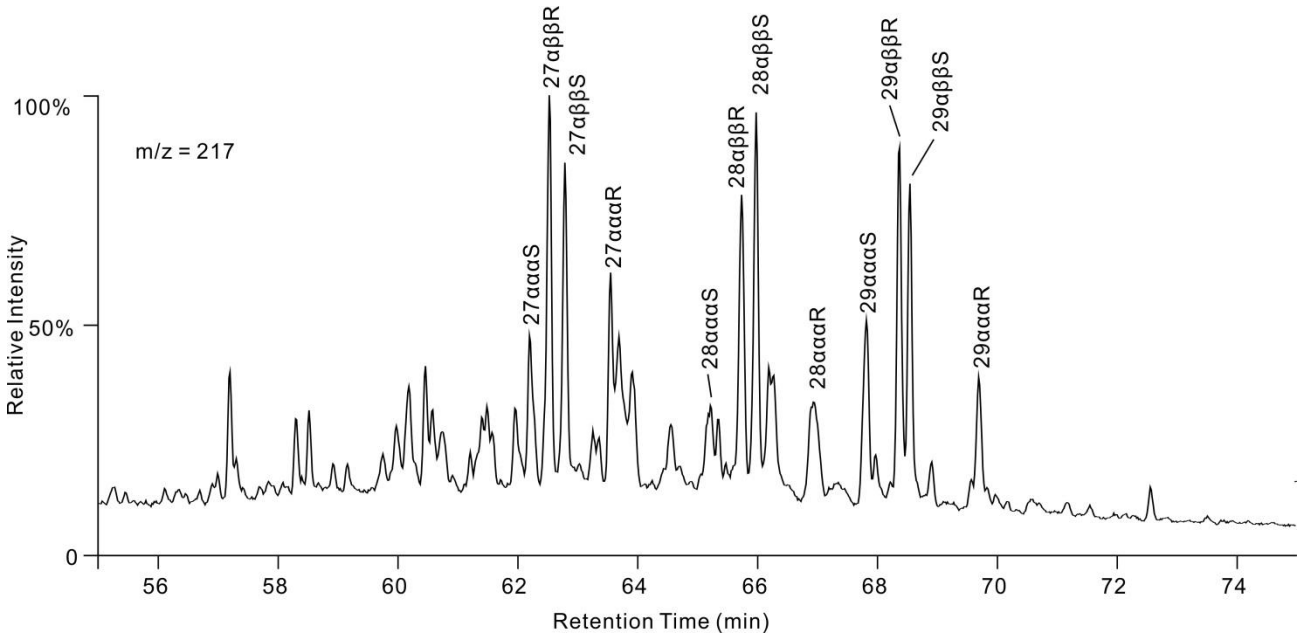


Fig. 8. GC-derived pristane/*n*-C_{17} versus phytane/*n*-C_{18} for depositional environment typing.

C_{27}, C_{28} and C_{29} sterol stereoisomers have been identified in m/z = 217 mass chromatograms of saturated fractions (Fig. 9). It has been known for almost forty years that distributions of C_{27}, C_{28} and C_{29} sterol homologues on a ternary diagram can be used to differentiate ecosystems (Huang and Meinschein, 1979), yet the discrimination of source depositional settings has been seldom effective because there is much overlap on this figure. when plotting a composite of data for oils from various source rock depositional environments (Moldowan et al., 1985). Therefore, the principal use of C_{27}-C_{28}-C_{29} sterane ternary is to distinguish groups of crude oils, reservoir and source rock extracts, and to correlate the relationships between crude oils and source bitumens (Peters et al., 2000). In accordance with the observation from isoprenoid biomarkers presented above, the distributions of steranes distinguish crude oil and reservoir extract populations as being genetically and intimately correlated, whereas source bitumens scatter significantly again from a predominance of C_{27} to a predominance of C_{29} sterol homologs (Fig. 10). Inferred ecosystem types shown on the background only serves as a reference rather than indicators, because this concept proposed by Huang and Meinschein (1979) have met only limited success when applying it to steranes in source rocks and crude oils (Mackenzie et al., 1983; Moldowan et al., 1985).



379

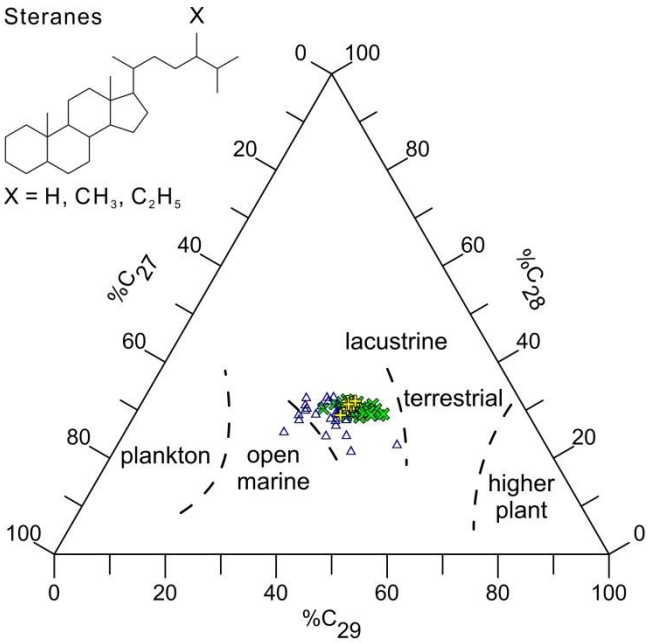
380 **Fig. 9.** Mass fragmentogram (m/z = 217) of the saturated fraction from a crude oil sample (R104731) showing typical distributions of C_{27}, C_{28} and C_{29}

381 steranes used in this study. C_{27} sterane consists of four stereoisomers, i.e., (20S)-5α(H),14α(H),17α(H)-cholestane, (20R)-5α(H),14β(H),17β(H)-cholestane,

382 (20S)-5α(H),14β(H),17β(H)-cholestane and (20R)-5α(H),14α(H),17α(H)-cholestane. C_{28} and C_{29} are methyl-cholestane and ethyl-cholestane, respectively,

383 consisting of corresponding stereoisomers as shown for C_{27} sterane.

384



385

Fig. 10. Ternary diagram of C_{27}, C_{28} and C_{29} sterane compositions for crude oils, reservoir and source rock extracts based on high-resolution biomarker analysis. Chemical structures are shown at left. The corners of the triangle represent the relative percentage of the corresponding sterane homologs. Inferred types of ecosystems are shown on the background modified from Huang and Meinschein (1979). Symbol legend refers to Fig. 8.

5. Discussions

5.1. Thermal maturity at time of expulsion

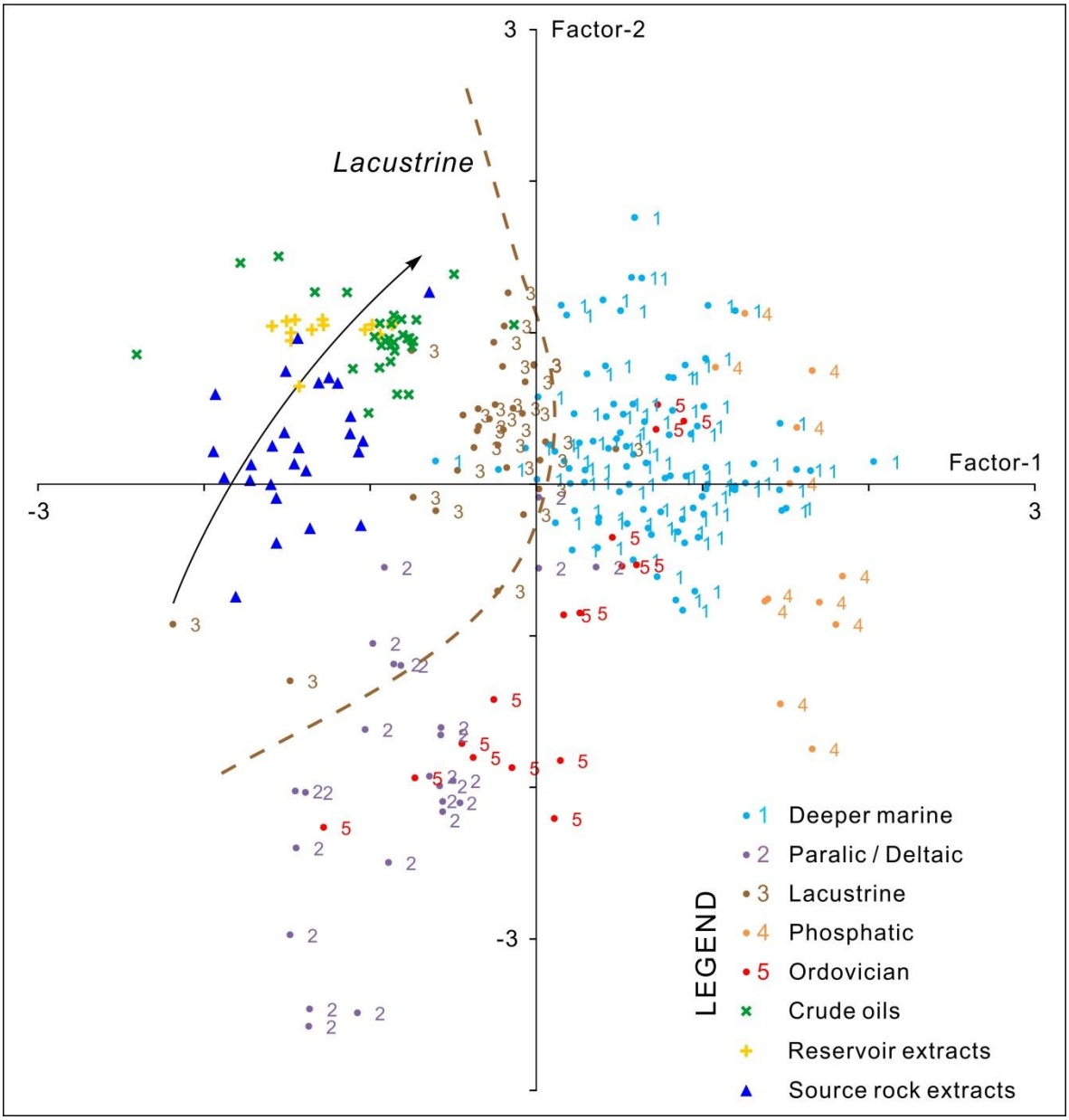
Curiale (2008) has argued there are the limitations in applying an oil-source correlation based solely on geochemical information, as this information can be affected by the choice of inappropriate samples with differing thermal maturity levels. As shown above, similar values of Methyldibenzothiophene Ratios (MDR1 and MDR2,3) as well as a biomarker parameter based on relative stability of C_{27} hopanes reflect that the samples were generated at similar maturity levels. Although the MDR4 ratio displays a scattered change, it shows a higher maturity for source rocks extracts than for the other two sample groups (Fig. 6). The same trend could also be acquired through all the remaining parameters. Zhang et al. (2013) have reported a similar finding as well using the Methylphenanthrene Index and Dimethyldibenzothiophene Index to estimate maturity levels in the Ordos Basin. This phenomenon of crude oils expelled from source rock or retained in reservoir being less mature than bitumen extracts is likely caused by differential maturation. After oils had been expelled into shallower carrier system from the source rock, those source rocks continued to be subjected to thermal evolution whereas the maturation of the expelled oils had ceased or slowed down considerably (Zhao et al., 1996; Yang et al., 2005; Ren et al., 2007). Therefore, benefiting from the local tectonic evolution during the Late Triassic, having relatively continuous thermal evolution, samples from the petroliferous Yanchang Formation alleviates the maturity differences issue proposed by Curiale (2008) and make a reasonable following discussion on source-reservoir-oil correlation and depositional discrimination.

5.2. Sample grouping in Q-mode factor analysis

Using the same factors as in Fig. 5, Fig. 11 shows the grouping distributions for the combined data set used in the aforementioned factor analysis. Groups 1 through 5 samples represent crude oils with known or deduced source rock depositional environment, and samples from Yanchang Formation data set are scattered in the lacustrine region confined by group 3 crude oils. Although the distribution pattern of the Q-mode factor analysis in this study is not specifically to the

same as that described by Zumberge (1987), the relative grouping configuration for each group reveals the same discrimination for different origin and source environment in both studies. For example, there are some overlaps between different groups of environmentally related crude oils; different source rock domains are grouping together to define some specific regions; each group has the same associated relationship with surrounding groups. These different distribution patterns result from the contribution of Yanchang Formation data participating in the whole factor analysis procedure, and the alternation of each variable on factor loadings in R-mode factor analysis. The good discrimination for each group suggests that the RQ-mode factor analysis of a combined sample set introduced in this study produces a good grouping pattern which is, therefore, entirely consistent with the results of published research and is therefore valid for further discussion in more detail.

The terpane biomarker distributions in crude oils or rock extracts can be used to estimate the nature of their respective depositional environment, and the origin of the parent organic matter. The aforementioned overlaps among five groups of crude oils owe their origins to the mixed nature of oils generated from different source facies or from a source rock but inheriting chemical features from different kinds of organic matter precursors. Zumberge (1987) has presented a detailed report on the classification of these five groups, thus no more discussion will be included here but further interpretation will be focused on the Yanchang Formation data set. Three distinctive groups can be discriminated in the Q-mode factor analysis (Fig. 11), though there is some degree of overlap, especially between the crude oils and reservoir extracts. The deltaic sandstone associations in the Chang 6 Member (Fig. 2) are the reservoir that is holding the crude oil produced from the studied field. An arrow describes the migration path for oils presumably from an identical source-reservoir system. This trend is not obscured by the method of RQ-mode factor analysis, though the variables have been normalized to one another, and a degree of freedom is lost during multivariate analysis (Davis, 1986).



433

434 **Fig. 11.** Q-mode factor analysis showing the interrelationship of the crude oil samples (groups 1, 2, 3, 4 and 5) with known source rock features from
435 multi-references as well as the crude oils, reservoir and source rock extracts from this study in factor- 1 and 2 space. These are the same factors as in Fig.
436 5 and account for 45 percent of the variation in the data. Zumberge (1987) has collected sample groups 1, 2, 3, 4 and 5 from various scholars and
437 materials. Dotted line showing a lacustrine environment confined by group 3 crude oils.

438

439 The detailed interrelationship between variables and samples for groups 1 to 5 has been fully discussed by Zumberge
440 (1987), thus only lacustrine samples will be referred to here. The combined samples from group 3 and Yanchang
441 Formation are characterized by the C_{21} tricyclic terpane and higher molecular weight pentacyclic terpanes (C_{26}, C_{27}, C_{28} and

442 C_{30}) association. Gammacerane exerts a strong contribution on ascertaining lacustrine environment in Zumberge (1987)'s
443 study but it doesn't work here. The Yanchang Formation data incorporated into this factor analysis may dilute the
444 concentration of gammacerane and weaken its significance, because the relative abundance of gammacerane in Yanchang
445 Formation samples is quite low (mean 2.19%) compared to that in Zumberge (1987)'s data (mean 12.40%). Commonly, the
446 gammacerane is highly specific for water-column stratification in lacustrine source rocks (Mann et al., 1987; Damsté and
447 De Leeuw, 1995; Chen and Summons, 2001), especially resulting from hypersalinity (Peters et al., 2007). The low
448 concentration of gammacerane is also supported by the reconstructed water chemistry composition of Ordos Lake which
449 was predominantly freshwater without salinity stratification at the time of deposition (Pan et al., 2016).

450 For the Yanchang samples, different distributive areas of crude oils, sandstone extracts and source rock bitumens are
451 controlled by different terpane associations (Fig. 11). Generated crude oils could be distinguished by relatively greater
452 quantities of C_{26} and C_{28} terpanes, whereas their carrier extracts contains more C_{21} and C_{30} concentrations. Although these
453 two kinds of samples seems to be governed by different terpane associations, they actually share genetic a relationship
454 because the sandstone extracts represent the reservoir (Chang 6 Member) from which the crude oils have been exploited.
455 Moreover, the source rocks bitumens display a greater scatter, and contain relatively abundant amounts of C_{27} pentacyclic
456 terpane.

457 *5.3. Geochemical characterization*

458 Sterane carbon-number distributions (Fig. 10) support the hypothesis that the studied crude oil samples and reservoir
459 extracts share a genetic relationship, as revealed by terpane-derived Q-mode factor analysis (Fig. 11). A composite of
460 maturity-related parameters, biomarkers and non-biomarkers, are indicative of a relatively consistent maturity level
461 between crude oils and reservoir extracts (Fig. 6). The *n*-alkane distributions of reservoir extracts derived from saturate
462 fraction gas chromatography are similar to crude oils (Figs. 7a and c), as are the distributions of the total ion current from
463 GC-MS identification (Figs. 7b and d). However, crude oils characteristically contain higher relative concentrations of
464 tricyclic terpanes compared to reservoir extracts, but much lower concentrations of hopanes and gammacerane (Fig. 12).
465 It is widely recognized that most commercial oil accumulations are largely charged and sourced from a mixed-input source
466 system (Seifert, 1979; Peters et al., 1989; Dzou et al., 1999; Chen et al., 2003a; Chen et al., 2003b), and so is the case for
467 the oils from the Yanchang Formation. This mixing effect may suppress the differentiation attempt using the C_{27}-C_{28}-C_{29}
468 steranes plotting approach but could be obviously overcome in the combined RQ-mode factor analysis based on a set of

terpane biomarker parameters. It also provides a hint on the separation between crude oil and source bitumen populations.

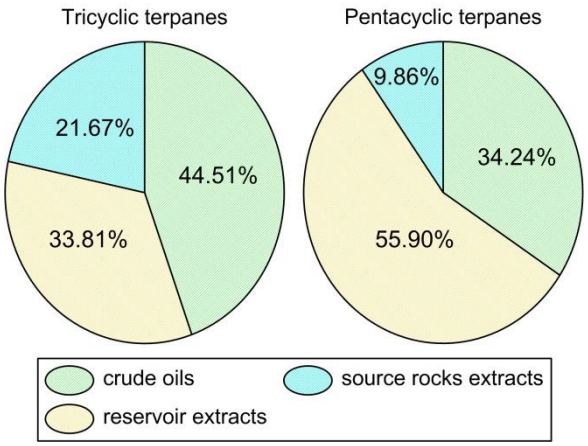


Fig. 12. Pie charts showing relative percentage of crude oils, source rock extracts and reservoir extracts in tricyclic and pentacyclic terpanes.

When considering oil-source correlation, genetic relationships seem to be more complicated as far as Q-mode factor analysis is concerned (Fig. 11). The investigated crude oils and source bitumens appear to be less correlated with each other than the reservoir extracts and crude oils are. The lower molecular weight C_{21} tricyclic terpane as well as the combination of Ts and Tm hopanes exert more primary control on source bitumens, whereas much heavier molecular weight C_{26} and C_{28} terpanes seem to be crucial determinations on crude oils (Figs. 5 and 11). From the aspects of chain length distribution and aromatic compound composition, extracts from source rocks show much higher concentrations of lower molecular weight *n*-alkanes, and have chain lengths shorter than *n*-C_{31} (Figs. 7g and h). Some source rocks are deposited under a typical reducing environment whereas others are linked with dysoxic or even intensive oxidizing conditions (Fig. 8). For example, the pristane is notably high in sample R101409 and so as the Pr/Ph value (4.57), both indicating a source rock subjected to an oxidizing process. Therefore, neither molecular distributions nor biomarker compositions show a simple and direct correlation between oils and source rocks. Some samples from source bitumens, however, apparently show a genetic homogeneity with crude oils. For example, no evidence of biodegradation is documented for source rocks that are similar to the aforementioned crude oils and reservoir extracts. The chain length distribution from a source rock extracts (i.e., R104766, Fig. 7e) resembles those from crude oils and reservoir, irrespective of shorter carbon numbers only reaching 30.

489 C_{27}, C_{28} and C_{29} steranes on a ternary diagram could also support the scenario of source-oil correlation discussed above,
490 i.e., crude oils have a close genetic relationship with reservoir extracts, while source bitumens seem to be separate from
491 the other two (Fig. 10). The reason for using C_{27}-C_{28}-C_{29} steroids as indicators for distinguishing sample groups is that the
492 distributions of these sterols do not appear to be extensively altered by catabolic or chemical processes (Huang and
493 Meinschein, 1979), and the plot locations on this ternary diagram also do not change significantly throughout the oil-
494 generative window (Peters et al., 1989; Peters et al., 2007). Plots of source rock extracts extend widely through C_{27} end
495 member to C_{29}. Commonly, phytoplankton is the dominant source of C_{27} sterol whereas the dominant input of C_{29} sterol is
496 related to higher plants (Huang and Meinschein, 1979). One outlier (R101409) is close to the C_{29} apex, and it seems to be
497 derived from terrigenous Type III organic matter. This hypothesis could also be supported by the same outlier from other
498 independent isoprenoid evidence in Fig. 8. In contrast, the vast majority of source rock samples are clustered close to the
499 C_{27} apex, which reflects a compositional origin from phytoplankton. Therefore, the source mixture is one possibility for
500 explaining why source bitumens seem to be genetically unrelated to the combination of crude oils and reservoir extracts.
501 This phenomenon suggests that the source system is predominated by the origin of aquatic algal, though the terrigenous
502 higher plants input cannot be excluded.

503 Our previous work focused on different types of organofacies of the Yanchang Formation shales denotes a sub-member
504 unit, Chang 7-2 shale, is composed of Type III basinal margin organic matter and appears to be dominated by short to
505 moderate chain lengths. The remaining two units, Chang 7-1 and Chang 7-3, are typical lacustrine shales enriched in long
506 to moderate alkyl chains dominated by algal organic facies (Pan et al., 2016). When the generated oils are expelled from
507 the host source rocks, those from Chang 7-2 are enriched in C_{29} sterols and form a mixed oil accumulation nearby with oils
508 dominated by C_{27} sterols from adjacent Chang 7-1 and Chang 7-3. The mixing can occur along migrating pathways or in the
509 traps after reservoiring. Further consecutive work will be involved in testifying this hypothesis and determining the mixing
510 time for specific sterane compounds along the insights enlightened from this RQ-mode factor analysis.

511 ## 6. Conclusions

512 Three Methyldibenzothiophene Ratios (MDR1, MDR2,3 and MDR4) in conjunction with the ratio of Ts/(Ts + Tm) have
513 been used to determine the thermal maturity levels of the studied samples. Consistent maturity levels estimated by these
514 parameters through the investigated Yanchang Formation minimise complications of maturity differences, and lead to a
515 sound discussion on source-reservoir-oil correlation and depositional environment typing. Through the combined RQ-

mode factor analysis based on sixteen terpane biomarkers, the relationships between both statistical variables and geological samples can be visualized in the same factor space providing assistance for geochemical interpretation. R-mode factor analysis shows that the C_{21} tricyclic terpane coupled with some higher molecular weight pentacyclic terpanes (C_{26}, C_{27}, C_{28} and C_{30}) exert a primary control upon lacustrine Yanchang Formation samples. More specifically, C_{26} and C_{28} terpanes are mainly responsible for distinguishing crude oil groups, while an association of abundant C_{21} and C_{30} terpanes characterizes their carrier extracts. C_{27} pentacyclic terpane has been found in relatively abundant amounts when examining source rock extracts. Besides, obvious interrelationships among crude oils, reservoir extracts and source bitumens have been identified with the inclusion of published data for 216 samples. All studied sample plot in a lacustrine facies region, and distributing along a source-reservoir-oil migration pathway according to the Q-mode factor analysis.

Furthermore, molecular composition adds more detailed information to Q-mode factor analysis on source-reservoir-oil correlation. No matter how many geochemical categories are involving in reservoir-oil correlation, for example, the molecular composition, chain length distribution, biomarker characterization or factor analysis grouping, crude oils and reservoir extracts from Yanchang Formation are always displaying close genetic interrelationships or consistent genetic similarities within each group. Most of the reservoir-oil samples originate from Type II kerogen or mixed Type II and III organic matter deposited under reducing conditions, and display a tight distribution using the relative concentrations of C_{27}-C_{28}-C_{29} steranes. Nevertheless, when including source rocks into the reservoir-oil system, depositional environment varies notably from typically anoxic to mixed dysoxic, even oxic conditions.

Utilizing the raw data or geochemical ratios is effective when using results from a single laboratory; however, interlaboratory differences can sometimes cause problems, and cannot be neglected. Thus, statistical approach presented in this study has generated a factor space containing several dimensionless variables, and provides a possibility to overcome the experimental incomparability caused by that issue. Consequently, the results from Zumberge could be used directly. For our case, the investigated source rocks could be the effective source for local crude oils, yet other source types enriched in terrigenous organic matter cannot be excluded. Combined RQ-mode factor analysis thus, could provide some useful clues, and should be accepted as an additional tool to source-reservoir-oil correlation which is based on traditional geochemical protocols.

Acknowledgement

We would like to express our gratitude to two anonymous reviewers for their comments and suggestions, and the editors for their help and advice. Special thanks to Senhu Lin, Songtao Wu and Dr. Jingwei Cui for the assistance of sample collection. We would appreciate the technicians from the laboratory of RIPED as well for their technical support. Our thanks are also extended to the Changqing Oilfield Company of PetroChina for support in materials collecting and permission to publish. This study is financially supported by the National Basic Research Program of China (No. 2014CB239000) and the Chinese Major Scientific and Technological Program (No. 2016ZX05406-006). The China Scholarship Council (CSC No. 201406010066) is acknowledged for financing Songqi Pan studying at GFZ German Research Centre for Geosciences.

References

Alexander, R., Kagi, R., Sheppard, P., 1984. 1,8-Dimethylnaphthalene as an indicator of petroleum maturity. Nature 308, 442-443.

Alexander, R., Kagi, R.I., Rowland, S.J., Sheppard, P.N., Chirila, T.V., 1985. The effects of thermal maturity on distributions of dimethylnaphthalenes and trimethylnaphthalenes in some Ancient sediments and petroleums. Geochimica Et Cosmochimica Acta 49, 385-395.

Bennett, B., Larter, S.R., 2000. Quantitative separation of aliphatic and aromatic hydrocarbons using silver ion-silica solid-phase extraction. Analytical Chemistry 72, 1039-1044.

Brooks, J.D., Gould, K., Smith, J.W., 1969. Isoprenoid Hydrocarbons in Coal and Petroleum. Nature 222, 257-259.

Chen, J., Deng, C., Liang, D., Wang, X., Zhong, N., Song, F., Shi, X., Jin, T., Xiang, S., 2003a. Mixed oils derived from multiple source rocks in the Cainan oilfield, Junggar Basin, Northwest China. Part II: artificial mixing experiments on typical crude oils and quantitative oil-source correlation. Organic Geochemistry 34, 911-930.

Chen, J., Liang, D., Wang, X., Zhong, N., Song, F., Deng, C., Shi, X., Jin, T., Xiang, S., 2003b. Mixed oils derived from multiple source rocks in the Cainan oilfield, Junggar Basin, Northwest China. Part I: genetic potential of source rocks, features of biomarkers and oil sources of typical crude oils. Organic Geochemistry 34, 889-909.

Chen, J., Summons, R.E., 2001. Complex patterns of steroidal biomarkers in Tertiary lacustrine sediments of the Biyang Basin, China. Organic Geochemistry 32, 115-126.

Chen, Q.H., Li, W.H., Gao, Y.X., Guo, Y.Q., Feng, J.P., Zhang, D.F., Cao, H.X., Liang, J.W., 2007. The deep-lake deposit in the upper Triassic Yanchang formation in Ordos Basin, China and its significance for oil-gas accumulation. Science in China Series D-Earth Sciences 50, 47-58.

Curiale, J.A., 1993. Oil to Source Rock Correlation, in: Engel, M.H., Macko, S.A. (Eds.), Organic Geochemistry: Principles and Applications. Springer US, Boston, MA, pp. 473-490.

Curiale, J.A., 1994. Correlation of Oils and Source Rocks--A Conceptual and Historical Perspective, in: Magoon, L.B., Dow, W.G. (Eds.), The petroleum system - from source to trap. AAPG Memoir 60, Tulsa, Oklahoma, pp. 251-260.

Curiale, J.A., 2008. Oil–source rock correlations – Limitations and recommendations. Organic Geochemistry 39, 1150-1161.

Damsté, J.S.S., De Leeuw, J.W., 1995. Comments on “Biomarkers or not biomarkers. A new hypothesis for the origin of pristane involving derivation from methyltrimethyltridecylchromans (MTTCs) formed during

581 diagenesis from chlorophyll and alkylphenols" from M. Li, S. R. Larter, P. Taylor, D. M. Jones, B. Bowler and
582 M. Bjorøy. Organic Geochemistry 23, 1085-1087.
583 Davis, J.C., 1986. Statistics and data analysis in geology, 2nd ed. Wiley, New York.
584 Didyk, B.M., 1978. Organic geochemical indicators of palaeoenvnonmental conditions of sedimentation. Nature
585 272, 216-222.
586 Duan, Y., Wang, C.Y., Zheng, C.Y., Wu, B.X., Zheng, G.D., 2008. Geochemical study of crude oils from the Xifeng
587 oilfield of the Ordos basin, China. Journal of Asian Earth Sciences 31, 341-356.
588 Dzou, L.I., Holba, A.G., Ramón, J.C., Moldowan, J.M., Zinniker, D., 1999. Application of new diterpane
589 biomarkers to source, biodegradation and mixing effects on Central Llanos Basin oils, Colombia. Organic
590 Geochemistry 30, 515-534.
591 Hanson, A.D., Ritts, B.D., Moldowan, J.M., 2007. Organic geochemistry of oil and source rock strata of the
592 Ordos Basin, north-central China. AAPG Bulletin 91, 1273-1293.
593 Hartough, H.D., Meisel, S.L., 1954. Compounds with condensed thiophene rings. Interscience Publishers, Inc.,
594 New York.
595 Hase, A., Hites, R.A., 1976. On the origin of polycyclic aromatic hydrocarbons in recent sediments: biosynthesis
596 by anaerobic bacteria. Geochimica Et Cosmochimica Acta 40, 1141-1143.
597 Hill, R.J., Jarvie, D.M., Zumberge, J., Henry, M., Pollastro, R.M., 2007. Oil and gas geochemistry and petroleum
598 systems of the Fort Worth Basin. AAPG Bulletin 91, 445-473.
599 Horsfield, B., 1989. Practical criteria for classifying kerogens: Some observations from pyrolysis-gas
600 chromatography. Geochimica Et Cosmochimica Acta 53, 891-901.
601 Huang, H.P., Zhang, S.C., Su, J., 2015. Geochemistry of Tri- and Tetracyclic Terpanes in the Palaeozoic Oils from
602 the Tarim Basin, Northwest China. Energy & Fuels 29, 7014-7025.
603 Huang, W.-Y., Meinschein, W.G., 1979. Sterols as ecological indicators. Geochimica Et Cosmochimica Acta 43,
604 739-745.
605 Ji, L.M., He, C., Zhang, M.Z., Wu, Y.D., Li, X.B., 2016. Bicyclic alkanes in source rocks of the Triassic Yanchang
606 Formation in the Ordos Basin and their inconsistency in oil-source correlation. Marine and Petroleum
607 Geology 72, 359-373.
608 Ji, L.M., Wu, T., Li, L.T., 2007. Geochemical characteristics of kerogen in Yanchang Formation source rocks,
609 Xifeng area, Ordos Basin. Petroleum Exploration and Development 34, 424-428.
610 Jia, C., Chi, Y., 2004. Resource petential and exploration techniques of stratigraphic and subtle reservoirs in
611 China. Petroleum Science 1, 1-12.
612 Jiao, Y., Wu, L., Wang, M., Xu, Z., 2005. Forecasting the occurrence of sandstone-type uranium deposits by
613 spatial analysis: An example from the northeastern Ordos Basin, China, in: Mao, J., Bierlein, F.P. (Eds.),
614 Meeting the Global Challenge: Proceedings of the Eighth Biennial SGA Meeting. Springer, Berlin, pp. 273-275.
615 Jones, R.W., 1987. Organic facies, Advances in petroleum geochemistry, v. 2. Academic Press, London, pp. 1-90.
616 Klovan, J.E., 1975. R- and Q-Mode Factor Analysis, in: McCammon, R.B. (Ed.), Concepts in Geostatistics.
617 Springer, Berlin, Heidelberg, pp. 21-69.
618 Laflamme, R.E., Hites, R.A., 1978. The global distribution of polycyclic aromatic hydrocarbons in recent
619 sediments. Geochimica Et Cosmochimica Acta 42, 289-303.
620 Li, G., Lu, M., 2002. Atlas of Chinese petroliferous basins (in Chinese with English abstract). Petroleum Industry
621 Press, Beijing.
622 Li, Q., Yang, Z., Zou, C., 2015. Shale Oil Geological Characteristics and Resource Potential of Mesozoic Yanchang
623 Formation in Ordos Basin, North-Central China. Acta Geologica Sinica - English Edition 89, 257-257.
624 Liu, C.Y., Zhao, H.G., Zhao, J.F., Wang, J.Q., Zhang, D.D., Yang, M.H., 2008. Temporo-spatial Coordinates of
625 Evolution of the Ordos Basin and Its Mineralization Responses. Acta Geologica Sinica-English Edition 82,
626 1229-1243.
627 Liu, S., Yang, S., 2000. Upper Triassic-Jurassic sequence stratigraphy and its structural controls in the western
628 Ordos Basin, China. Basin Research 12, 1-18.
629 Mackenzie, A.S., 1984. Applications of biological markers in petroleum geochemistry. Advances in petroleum
630 geochemistry 1, 1-210.

631 Mackenzie, A.S., Disko, U., Rullkötter, J., 1983. Determination of hydrocarbon distributions in oils and sediment
632 extracts by gas chromatography—high resolution mass spectrometry. Organic Geochemistry 5, 57-63.
633 Mann, A.L., Goodwin, N.S., Lowe, S., 1987. Geochemical characteristics of lacustrine source rocks: a combined
634 palynological/molecular study of a Tertiary sequence from offshore China, Proceedings of the Indonesian
635 Petroleum Association, Sixteenth Annual Convention. Indonesian Petroleum Association, Jakarta, Indonesia,
636 pp. 241-258.
637 Moldowan, J.M., Seifert, W.K., Gallegos, E.J., 1985. Relationship between petroleum composition and
638 depositional environment of petroleum source rocks. AAPG Bulletin 69, 1255-1268.
639 Pan, S.Q., Horsfield, B., Zou, C.N., Yang, Z., 2016. Upper Permian Junggar and Upper Triassic Ordos lacustrine
640 source rocks in Northwest and Central China: Organic geochemistry, petroleum potential and predicted
641 organofacies. International Journal of Coal Geology 158, 90-106.
642 Peters, K.E., Fraser, T.H., Amris, W., Rustanto, B., Hermanto, E., 1999. Geochemistry of crude oils from eastern
643 Indonesia. AAPG Bulletin 83, 1927-1942.
644 Peters, K.E., Moldowan, J.M., Driscole, A.R., Demaison, G.J., 1989. Origin of Beatrice oil by co-sourcing from
645 Devonian and Middle Jurassic source rocks, inner Moray Firth, United Kingdom. AAPG Bulletin 73, 454-471.
646 Peters, K.E., Snedden, J.W., Sulaeman, A., Sarg, J.F., Enrico, R.J., 2000. A New Geochemical-Sequence
647 Stratigraphic Model for the Mahakam Delta and Makassar Slope, Kalimantan, Indonesia. AAPG Bulletin 84,
648 12-44.
649 Peters, K.E., Walters, C.C., Moldowan, J.M., 2007. The biomarker guide volume 2: biomarkers and isotopes in
650 petroleum systems and earth history. Cambridge University Press.
651 Powell, T.G., 1988. Pristane/phytane ratio as environmental indicator. Nature 333, 604.
652 Powell, T.G., McKirdy, D.M., 1973. Relationship between ratio of pristane to phytane, crude oil composition
653 and geological environment in Australia. Nature Physical Science 243, 37-39.
654 Qiu, Z., Gong, Z., 1999. Petroleum exploration in China v. 2: Western Petroleum Province (in Chinese with
655 English abstract). Petroleum Industry Press, Beijing.
656 Radke, M., 1988. Application of aromatic compounds as maturity indicators in source rocks and crude oils.
657 Marine and Petroleum Geology 5, 224-236.
658 Radke, M., Leythaeuser, D., Teichmüller, M., 1984. Relationship between rank and composition of aromatic
659 hydrocarbons for coals of different origins. Organic Geochemistry 6, 423-430.
660 Radke, M., Welte, D.H., Willsch, H., 1982. Geochemical study on a well in the Western Canada Basin: relation of
661 the aromatic distribution pattern to maturity of organic matter. Geochimica Et Cosmochimica Acta 46, 1-10.
662 Radke, M., Welte, D.H., Willsch, H., 1986. Maturity parameters based on aromatic hydrocarbons: Influence of
663 the organic matter type. Organic Geochemistry 10, 51-63.
664 Ren, Z., Zhang, S., Gao, S., Cui, J., Xiao, Y., Xiao, H., 2007. Tectonic thermal history and its significance on the
665 formation of oil and gas accumulation and mineral deposit in Ordos Basin. Science in China Series D-Earth
666 Sciences 50, 27-38.
667 Seifert, W.K., 1979. Application of biological marker chemistry to petroleum exploration, Proceedings of the
668 10th World Petroleum Congress. Heyden & Son Ltd, London, pp. 425-440.
669 Seifert, W.K., Michael Moldowan, J., 1978. Applications of steranes, terpanes and monoaromatics to the
670 maturation, migration and source of crude oils. Geochimica Et Cosmochimica Acta 42, 77-95.
671 Seifert, W.K., Michael Moldowan, J., 1979. The effect of biodegradation on steranes and terpanes in crude oils.
672 Geochimica Et Cosmochimica Acta 43, 111-126.
673 Seifert, W.K., Moldowan, J.M., 1986. Use of biological markers in petroleum exploration, in: Johns, R.B. (Ed.),
674 Methods in geochemistry and geophysics. Elsevier, Amsterdam, pp. 261-290.
675 Soxhlet, F.v., 1879. Die gewichtsanalytische Bestimmung des Milchfettes. Polytechnisches Journal (Dingler's)
676 232, 461-465 (in German).
677 Sun, Z., Xie, Q., Yang, J., 1989. Ordos Basin—a typical example of an unstable cratonic interior superimposed
678 basin, in: Zhu, X. (Ed.), Chinese sedimentary basins. Elsevier, Amsterdam, pp. 63-75.
679 Tao, S., Wang, C., Du, J., Liu, L., Chen, Z., 2015. Geochemical application of tricyclic and tetracyclic terpanes
680 biomarkers in crude oils of NW China. Marine and Petroleum Geology 67, 460-467.

681 Temple, J.T., 1978. The use of factor analysis in geology. Journal of the International Association for
682 Mathematical Geology 10, 379-387.
683 Tissot, B., Califet-Debyser, Y., Deroo, G., Oudin, J.L., 1971. Origin and Evolution of Hydrocarbons in Early
684 Toarcian Shales, Paris Basin, France. AAPG Bulletin 55, 2177-2193.
685 Wakeham, S.G., Schaffner, C., Giger, W., 1980. Poly cyclic aromatic hydrocarbons in Recent lake sediments—II.
686 Compounds derived from biogenic precursors during early diagenesis. Geochimica Et Cosmochimica Acta 44,
687 415-429.
688 Walden, J., Smith, J.P., Dackombe, R.V., 1992. The use of simultaneous R- and Q-mode factor analysis as a tool
689 for assisting interpretation of mineral magnetic data. Mathematical Geology 24, 227-247.
690 Watson, M.P., Hayward, A.B., Parkinson, D.N., Zhang, Z.M., 1987. Plate tectonic history, basin development and
691 petroleum source rock deposition onshore China. Marine and Petroleum Geology 4, 205-225.
692 Yang, H., Deng, X.Q., 2013. Deposition of Yanchang Formation deep-water sandstone under the control of
693 tectonic events in the Ordos Basin. Petroleum Exploration and Development 40, 549-557.
694 Yang, H., Zhang, W., Wu, K., Li, S., Peng, P.a., Qin, Y., 2010. Uranium enrichment in lacustrine oil source rocks of
695 the Chang 7 member of the Yanchang Formation, Erdos Basin, China. Journal of Asian Earth Sciences 39,
696 285-293.
697 Yang, Y.T., Li, W., Ma, L., 2005. Tectonic and stratigraphic controls of hydrocarbon systems in the Ordos basin:
698 A multicycle cratonic basin in central China. AAPG Bulletin 89, 255-269.
699 Yang, Z., He, S., Guo, X.W., Li, Q.Y., Chen, Z.Y., Zhao, Y.C., 2016. Formation of low permeability reservoirs and
700 gas accumulation process in the Daniudi Gas Field, Northeast Ordos Basin, China. Marine and Petroleum
701 Geology 70, 222-236.
702 Yang, Z., Hou, L., Tao, S., Cui, J., Wu, S., Lin, S., Pan, S., 2015. Formation and “sweet area” evaluation of liquid-
703 rich hydrocarbons in shale strata. Petroleum Exploration and Development 42, 609-620.
704 Yu, J., Yang, Y., Du, J., 2010. Sedimentation during the transgression period in Late Triassic Yanchang Formation,
705 Ordos Basin. Petroleum Exploration and Development 37, 181-187.
706 Zhai, G., 1997. Petroleum geology of China. Petroleum Industry Press, Beijing.
707 Zhang, L., Li, M., Wang, Y., Yin, Q.-Z., Zhang, W., 2013. A novel molecular index for secondary oil migration
708 distance. Scientific Reports 3, 2487.
709 Zhang, S., Huang, H., 2005. Geochemistry of Palaeozoic marine petroleum from the Tarim Basin, NW China:
710 Part 1. Oil family classification. Organic Geochemistry 36, 1204-1214.
711 Zhang, S., Su, J., Wang, X., Zhu, G., Yang, H., Liu, K., Li, Z., 2011. Geochemistry of Palaeozoic marine petroleum
712 from the Tarim Basin, NW China: Part 3. Thermal cracking of liquid hydrocarbons and gas washing as the
713 major mechanisms for deep gas condensate accumulations. Organic Geochemistry 42, 1394-1410.
714 Zhao, M., Behr, H.-J., Ahrendt, H., Wemmer, K., Ren, Z., Zhao, Z., 1996. Thermal and tectonic history of the
715 Ordos Basin, China; evidence from apatite fission track analysis, vitrinite reflectance, and K-Ar dating. AAPG
716 Bulletin 80, 1110-1134.
717 Zhou, D., Chang, T., Davis, J.C., 1983. Dual extraction of R-mode and Q-mode factor solutions. Journal of the
718 International Association for Mathematical Geology 15, 581-606.
719 Zou, C.N., Tao, S.Z., Yuan, X.J., Zhu, R.K., Hou, L.H., Wang, L., Gao, X.H., Gong, Y.J., 2009. The formation
720 conditions and distribution characteristics of continuous petroleum accumulations (in Chinese with English
721 abstract). Acta Petrolei Sinica 30, 325-331.
722 Zou, C.N., Wang, L., Li, Y., Tao, S.Z., Hou, L.H., 2012. Deep-lacustrine transformation of sandy debrites into
723 turbidites, Upper Triassic, Central China. Sedimentary Geology 265–266, 143-155.
724 Zou, C.N., Yang, Z., Hou, L.H., Zhu, R.K., Cui, J.W., Wu, S.T., Lin, S.H., Guo, Q.L., Wang, S.J., Li, D.H., 2015.
725 Geological characteristics and “sweet area” evaluation for tight oil. Petroleum Science 12, 606-617.
726 Zou, C.N., Yang, Z., Tao, S.Z., Yuan, X., Zhu, R.K., Hou, L.H., Wu, S.T., Sun, L., Zhang, G.S., Bai, B., Wang, L., Gao,
727 X.H., Pang, Z.L., 2013. Continuous hydrocarbon accumulation over a large area as a distinguishing
728 characteristic of unconventional petroleum: The Ordos Basin, North-Central China. Earth-Science Reviews
729 126, 358-369.

730 Zumberge, J.E., 1987. Prediction of source rock characteristics based on terpane biomarkers in crude oils: A
731 multivariate statistical approach. Geochimica Et Cosmochimica Acta 51, 1625-1637.

732