Insights from principal component analysis applied to Py-GCMS study of Indian coals and their solvent extracted clean coal products

Abyansh Roy[1] · Heena Dhawan[1] · Sreedevi Upadhyayula[1] · Hariprasad Kodamana[1]

Abstract The present work aims at studying five Indian coals and their solvent extracted clean coal products using Py-GCMS analysis and correlating the characterization data using theoretical principal component analysis. The pyrolysis products of the original coals and the super clean coals were classified as mono-, di- and tri-aromatics, while other prominent products that were obtained included cycloalkanes, n-alkanes, and alkenes ranging from C_{10}–C_{29}. The principal component analysis is a dimensionality reduction technique that reduced the number of input variables in the characterization dataset and gave inferences on the relative composition of constituent compounds and functional groups and structural insights based on scores and loading plots which were consistent with the experimental observations. ATR-FTIR studies confirmed the reduced concentration of ash in the super clean coals and the presence of aromatics. The Py-GCMS data and the ATR-FTIR spectra led to the conclusion that the super clean coals behaved similarly for both coking and non-coking coals with high aromatic concentrations as compared to the raw coal. Neyveli lignite super clean coal was found to show some structural similarity with the original coals, whereas the other super clean coals showed structural similarity within themselves but not with their original coal samples confirming the selective action of the e,N solvent in solubilizing the polycondensed aromatic structures in the coal samples.

Keywords Indian coal · Super clean coals · Pyrolysis · PCA · ATR-FTIR · Py-GCMS

1 Introduction

Pyrolysis has been used to produce fuels and oils with high energy density and low volatility. There are various reactions involved in the pyrolysis of fossil fuels and biomass, such as aromatization, hydrogen transfer, oligomerization, and de-oxygenation. Many of these reactions produce aromatics. The product slate formed during pyrolysis includes liquid (oil), gases, water, char, and coke (Huber et al. 2006). A variety of analytical techniques such as Fluorescence Spectroscopy, Mass Spectrometry (MS), High-Performance Liquid Chromatography (HPLC), and Infrared Spectroscopy (IR) is typically used to analyze the complex mixture of hydrocarbons formed and their derived products (De Lira et al. 2010; Molina et al. 2010; Mullins 2010)

Solvent extraction of coal has been explored in the past (Van Krevelen 1993; Sharma and Dhawan 2018). The production of clean coal using various solvents has led to the recovery of Super Clean Coals (SCCs), Ultra-Clean Coals (UCCs), and Ultra-Super Clean Coals (U-SCCs) with negligible or no ash content. Hypercoal production is a recent development that uses two-ring aromatic solvents like tetralin and 1-Methylnaphthalene (1-MN) with which extraction of 70 wt% of bituminous coals is possible at elevated temperatures and high pressures (Okuyama et al. 2004; Masayuki et al. 2011). Solvent extraction of coals and the analysis of their clean products through pyrolysis

✉ Sreedevi Upadhyayula
   sreedevi@chemical.iitd.ac.in

✉ Hariprasad Kodamana
   kodamana@chemical.iitd.ac.in

[1] Department of Chemical Engineering, Indian Institute of Technology Delhi, Hauz Khas, New Delhi 110016, India

and Py-GCMS is an area gaining in importance (Fujitsuka et al. 2013; Yan et al. 2014; Zhang et al. 2016). Thermogravimetric analysis (TG) and Fourier Transform Infrared (FTIR) analysis of evolved products have also been successfully used to characterize heavy hydrocarbons, rocks, coals, biomass, waste materials, plastics, etc. FTIR provides insights into the pyrolysis of coals, including the kinetics of functional-group decomposition. The elements and functional groups of coal tar closely resemble coal and typical macromolecular debris of coal. Therefore, methods typically applied to coal analysis can reveal information about the composition of heavy hydrocarbon liquids.

Pyrolysis reactions are complex and require a better understanding through the study of reaction networks, kinetics and mechanisms. In this context, Py-GCMS is a particularly useful analytical method and can be used semiquantitatively to identify volatile products. The coal tar generated can prove to be a valuable product to derive much useful information. Coal tar contains chemical compounds that are very valuable commercially and industrially. These include aromatic compounds (benzene, toluene, xylene, naphthalene, anthracene, etc.), phenolic compounds (phenol, cresol, xylenol, cathecol, resorcinol, etc.), heterocyclic nitrogen compounds (pyridine, quinolone, isoquinoline, indole, etc.) and oxygen heterocyclic compounds (Fardhyanti 2015). The composition of coal tar varies with its origin and type of coal (for example lignite, bituminous coal, or anthracite) used in its production. Therefore, a more advanced data analysis technique could reveal more insights from the experimental data sets. In this work, we employ a combination of Principal Component Analysis (PCA) and Gaussian Mixture Modeling (GMM) for detailed data analysis and further categorization of these compounds.

PCA is a statistical analysis that helps to study and categorize many possibly interrelated parameters into a small number of uncorrelated variables called principal components. This technique is instrumental in extracting relevant inferences from complex data sets as it reduces a large set of variables into a smaller set without significant loss of the information. Indeed, PCA is a widely used technique for analyzing multivariate Py-GC/MS data (Meier et al. 2005; Schwarzinger 2005; Pattiya et al. 2010). PCA has also been used to characterize and classify several coals based on the differences in their properties, such as particle size, density, and ash contents (Niedoba 2014). A database for 44 brown coals has been studied extensively by Tesch and Otto (1995) to derive co-relations between several chemical, technological, and IR spectroscopic parameters. The principal component analysis helped explain the data and categorize the coals. Additional components were required to be included to derive more information from the data. Three Canadian coals, upgraded through evaporative drying up to 500 °C, were studied by Friesen and Ogunsola using PCA (Friesen and Ogunsola 1994). The lignite showed variation in its tar structure above 400 °C, whereas other two bituminous and sub-bituminous coals above 300 °C. No significant difference was found in the structure of the upgraded coals as compared to the original coals.

In the present study, PCA analysis was applied to identify, categorize and compare the data sets of five original coals (OCs) and their SCCs. PCA analysis was used to study the statistical nature and the relative concentrations of the aromatics, aliphatics, alkenes, etc., formed during the original coals' pyrolysis and was used to compare them with their solvent extracted clean products. GMM (Gaussian Mixture Model) algorithm was used along with PCA for identification and quantitative characterization of clusters in the plot. It facilitates better visualization of the spread of samples in space and statistical analysis of local features in the data.

2 Experiment

2.1 Samples and geological setting

Five Indian coal samples—Bhagabandh coal, Moonidih coal, Pandra coal, Bahula coal, Neyveli Lignite—were procured from NTPC power station, New Delhi, and the Ministry of Coal, Govt. of India. Neyveli lignite was procured from Neyveli Lignite Corporation. Tables S1 and S2 (in Supplementary data) show the proximate and ultimate analysis of the samples. The samples were dried and extracted with N-methyl pyrrolidone and a small amount of ethylenediamine. The procedure reported by Dhawan and Sharma (2019) was used to recover the SCCs. The extraction yields of the coals and the ash contents in the SCCs are given in Table S3 (in Supplementary data) (Fig. 1).

2.2 ATIR-FTIR spectra

The ATR-FTIR spectral analysis of the raw (OC), residual coal (left after the SCC extraction), and SCC was performed using Nicolet 6700 Infra-Red Spectrometer. ATR-FTIR spectra were recorded in the range 4000–600 cm^{-1} using a Ge crystal with a resolution of 4 cm^{-1} with 32 scans (Odeh 2015; Dhawan and Sharma 2019).

2.3 Py-GCMS analysis

Py-GCMS experiments were performed using a Pyroprobe Model 5200 (CDS Analytical, Inc.) connected to an Agilent 7890 GC with an Agilent 5975C MS detector. The
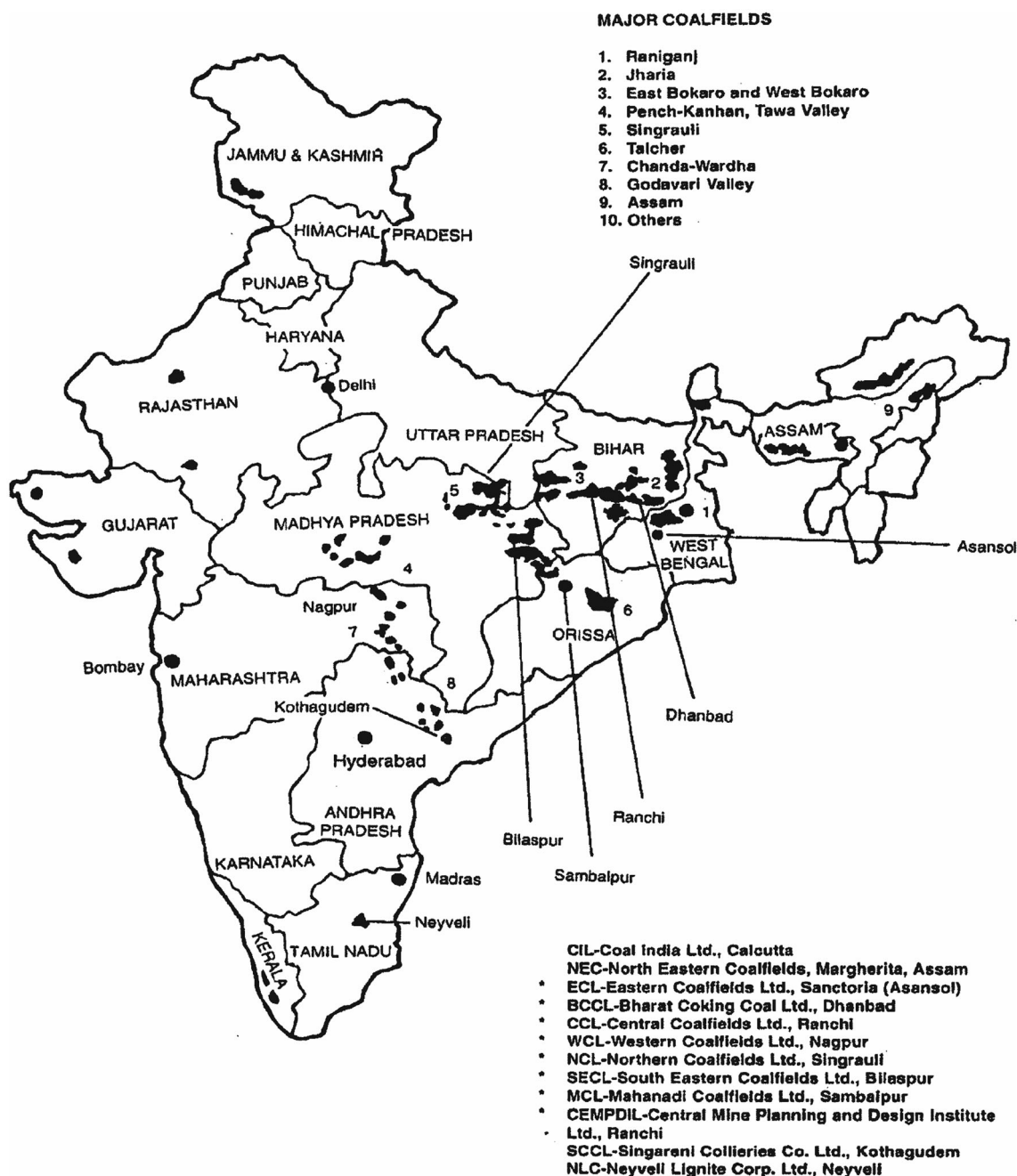
MAJOR COALFIELDS

1. Raniganj
2. Jharia
3. East Bokaro and West Bokaro
4. Pench-Kanhan, Tawa Valley
5. Singrauli
6. Talcher
7. Chanda-Wardha
8. Godavari Valley
9. Assam
10. Others

CIL-Coal India Ltd., Calcutta
NEC-North Eastern Coalfields, Margherita, Assam
• ECL-Eastern Coalfields Ltd., Sanctoria (Asansol)
• BCCL-Bharat Coking Coal Ltd., Dhanbad
• CCL-Central Coalfields Ltd., Ranchi
• WCL-Western Coalfields Ltd., Nagpur
• NCL-Northern Coalfields Ltd., Singrauli
• SECL-South Eastern Coalfields Ltd., Bilaspur
• MCL-Mahanadi Coalfields Ltd., Sambalpur
• CEMPDIL-Central Mine Planning and Design Institute
• Ltd., Ranchi
SCCL-Singareni Collieries Co. Ltd., Kothagudem
NLC-Neyveli Lignite Corp. Ltd., Neyveli

Fig. 1 Location of coal samples illustrated on map of India

pyroprobe was run in trap mode under He atmosphere. Pyrolysis was conducted at 650 °C (1000 °C/s heating rate) for 20 s. The valve oven and transfer lines were maintained at 325 °C. The column used in the GC was a DB1701 (60 m×0.25 mm×0.25 µm), and the temperature program was as follows: 45 °C for 3 min, ramp to 280 °C at 4 °C/min, and hold for 10 min. The flow rate was set to 1 mL/min using He as the carrier gas. The inlet and auxiliary lines were maintained at 300 °C, and the MS source was set at 70 eV. The GC-MS was calibrated for several phenolic compounds including phenol, 2-methoxyphenol, 2-methoxy-4-methylphenol, 2,6-dimethoxyphenol, vanillin, syringe aldehyde, and 2-methoxy-4-vinylphenol. Pyrolysis products were analyzed according to retention times and mass spectra data obtained from a NIST library.

One milligram of the ground (45–150 µm) coal (raw and SCCs) was analyzed in a quartz cell packed with quartz wool. Samples were heated to 100 °C for 10 s in the probe before analysis to remove any residual water. Prior to sample analysis, blank experiments were performed to

validate the cleanliness of the system. After sample analysis, methanol was run as a sample to remove any condensed products inside the pyroprobe. Methanol and blank experiments were repeated as necessary until the system was clean (Harman-Ware et al. 2015). All experiments were performed 3 times, and the average values were reported.

2.4 Data analysis

Principal Component Analysis was carried out using MATLAB's Statistics and Machine Learning Toolbox. After clubbing compounds into groups of aromatics and aliphatics, PCA was carried out with the sum of peak area percentages of compounds (belonging to that particular group) as variables/dimensions for the ten coal samples-five original (OC) and five supercleaned (SCCs). A particular type of coal, say Neyveli lignite, would have its own original and super cleaned variants. The principal components were obtained by Eigen Value decomposition of the data covariance matrix. The principal subspace is defined by the Eigen Vectors corresponding to the largest Eigen Values retained. Here, the data is presented with two principal components (PC1 and PC2), and thus, the effective dimensionality of data gets reduced to two. Scores and Loadings plot are referred together to analyze the results, with the former depicting the relative position of samples on the plane with PC1 and PC2 as abscissa and ordinate, respectively. GMM algorithm was used for determining cluster characteristics, i.e., cluster mean and covariance matrix. It was used to identify the clusters of super cleaned and original coals on the 2D plane with PC1 and PC2 as axes. Contour plots represent the stretch of clusters and the spread of data points inside them.

3 Results and discussions

3.1 ATR-FTIR spectra analysis

ATR-FTIR spectra for all coal samples verified the PCA analysis. Bands present between 1100 and 600 cm^{-1} correspond to mineral matter such as illite, montmonrollite, quartz, and aluminosilicates. The band specifically around 1030 cm^{-1} shows the presence of Si–O bonds (ash) in the system. The raw coal and residue clearly show the presence of ash, whereas the intensity of the band around 1030 cm^{-1} diminishes in the SCCs (Balachandran 2014). The aromatic absorption bands at 750 and 815 cm^{-1} in the raw coals and the SCCs are due to out-of-plane vibrations of one isolated and two adjacent C–H aromatic groups, respectively (Table 1). The spectra of the SCCs showed one such aromatic band at 750 cm^{-1}. Absorption bands at 1600 cm^{-1} in

SCCs are attributed to the presence of C=C stretching accentuated due to H-bond containing functional groups (Table 1) (Cooke et al. 1986). These aromatic bands were more intense in SCCs, showing effective action of the e,N solvent in coal cleaning (Sun et al. 2011). Pandra coal shows an increase in aromatics, especially triaromatics while, Bhagabandh shows an increase in diaromatics and triaromatics, which was also evident by the increase in the H$_{ar}$/H$_{ali}$ ratio.

3.2 Py-GCMS and PCA of OCs and SCCs

3.2.1 Py-GCMS

Py-GCMS studies were performed for the five coals and their solvent extracted SCCs. The basic characterization of these coals (Elemental, proximate analysis, TG curve) is already reported by the authors (experimentalists in this group, Dhawan et al.) in a previously published work (Sharma et al. 2019).

Each sample gave rise to more than 50 identifiable chemical compounds during pyrolysis. The compounds were categorized according to triaromatics, diaromatics, monoaromatics, cycloalkanes, *n*-alkanes, alkenes, and other structures based on matches with the NIST mass spectral library. The main aliphatic compounds present in the coal tar (i.e., pyrolysis vapor) were paraffin hydrocarbons, while the main aromatic components were monoaromatics such as benzene, toluene, o-xylene, phenols, and substituted phenols. Polycyclic aromatic hydrocarbons (PAHs) that were identified included naphthalene, substituted naphthalenes, phenanthrene, substituted phenanthrenes, anthracene, pyrene, etc., with naphthalene and its derivatives constituting the main constituents. The coal tar also contained some oxygen compounds, such as acids, and small amounts of aromatics containing nitrogen (Makan et al. 2017).

The mass spectra of the whole sample (OCs and SCCs) pyrolysis vapors show a complex series of ions ranging from $m/z = 100$ to $m/z = 450$, indicating the presence of polyaromatics such as fluoranthene and pyrene. Other significant products that were obtained through the analysis of the NL OC were methylpyrenes ($m/z = 216$), methyl phenol ($m/z = 220$), toluene ($m/z = 92$), and in NL SCC, high intensity of $m/z = 220$ and higher alkenes was observed.

An intense anthracene signal ($m/z = 178$) was observed for Pandra SCC while the product mixture from Pandra OC was mainly dominated by alkanes and alkenes (Fig. 2). An intense signal corresponding to o-xylene ($m/z = 106$) was observed in the case of both Moonidih OC and SCC (Fig. 3). Alkane peaks from C17 to C28 are observed at the end of the chromatogram. The narrow peaks (for C17 to

Table 1 Structural parameters deduced from the FTIR peak area measurements

Coal	H_{ar}/H_{al}	H_{ar} (wt%, 2700–3000 cm^{-1})	H_{ali} (wt%, 700–900 cm^{-1})	C_{ar} (wt%, 1600–1800 cm^{-1})	C_{ali} (wt%, 1375–1450 cm^{-1}, 2850–2990 cm^{-1})	C_{ar}/C_{al}	$-C=O$ (wt%) (2695–2830 cm^{-1})	CH_3 (deformation) (1300–1400 cm^{-1})	H_{ar}/C_{ar}
Bhagabandh OC	1.34	8.03	5.99	6.23	6.20	1.00	4.32	2.95	1.28
Bhagabandh SCC	1.55	9.29	5.98	6.82	5.73	1.19	3.78	2.55	1.36
Bahula OC	1.35	7.95	5.87	5.71	6.21	0.92	4.09	2.98	1.39
Bahula SCC	1.59	9.23	5.77	5.93	5.98	0.99	3.93	2.63	1.55
Moonidih OC	1.46	8.45	5.76	6.12	5.97	1.02	3.84	3.65	1.38
Moonidih SCC	1.66	9.71	5.83	5.75	6.21	0.93	3.90	3.07	1.68
Pandra OC	1.43	9.46	6.59	6.90	7.72	0.89	4.66	14.34	1.37
Pandra SCC	1.58	9.50	5.99	5.79	7.28	0.79	3.84	3.00	1.64
NL OC	1.53	8.76	5.74	5.85	5.93	0.98	3.54	3.00	1.50
NL SCC	1.47	8.82	5.97	6.30	7.01	0.89	3.74	2.95	1.4
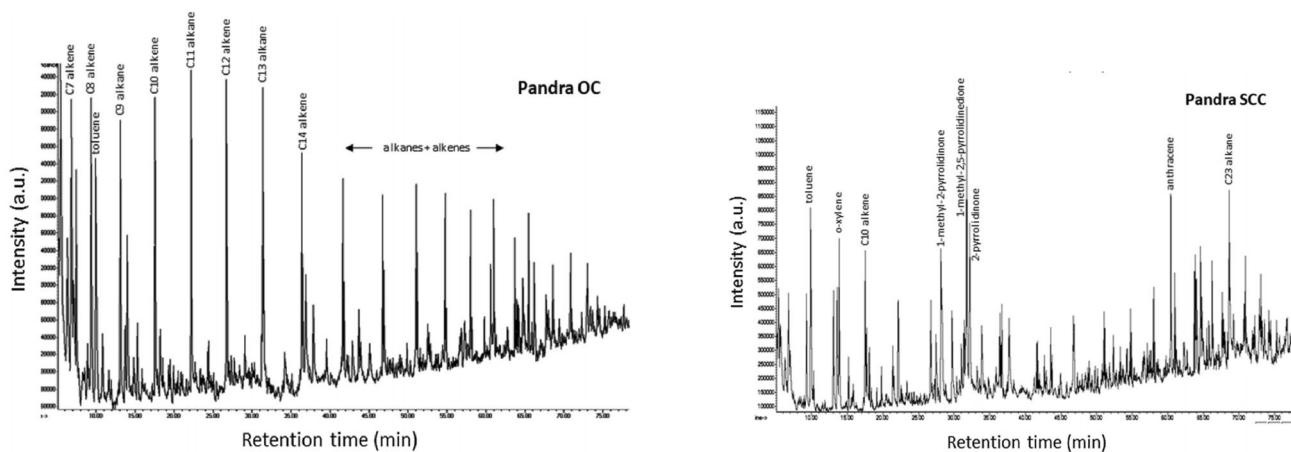
Fig. 2 Pyrograms of Pandra **a** OC and **b** SCC (Sharma et al. 2019)
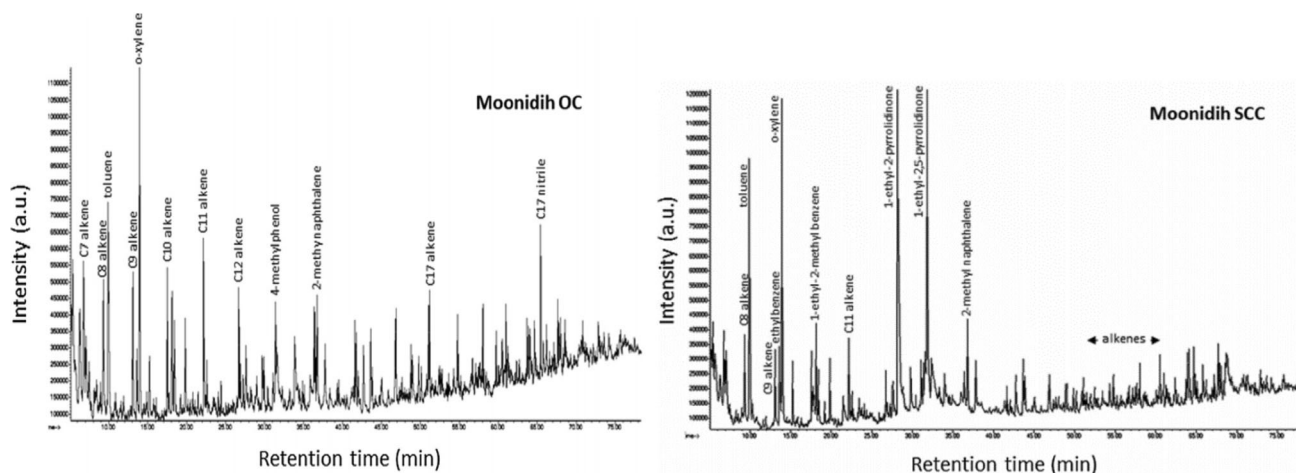
Fig. 3 Pyrograms of Moonidih **a** OC and **b** SCC (Sharma et al. 2019)

C28 alkanes and alkenes) vary from coal to coal, showing higher intensity in the case of high ash Pandra coal and low intensity in the case of low ash Bhagabandh and Moonidih coals (Figs. 2, 3). Toluene was found to be a major product from Bahula OC. Indeed, in general, the OCs were found to be good sources of monoaromatics such as benzene, toluene, and o-xylene (Islas et al. 2000).

The average aromatic content in the Moonidih OC was found to be marginally higher than the corresponding SCC (Fig. 3). In the case of the other four coals studied—Pandra coal (high ash non-coking coal), Bahula coal (high ash non-coking coal), Bhagabandh coal (low ash coking coal), and Neyveli lignite (low ash lignite), the average content of the aromatics in SCCs as compared to OCs is higher than the aliphatics. Aromatics larger than naphthalene could have been formed from char or other non-volatile products during the pyrolysis process or through numerous reactions involving alkylated derivatives (Sharma et al. 2019).

As explained by Radenovic (2006), the pyrolysis conditions result in the formation of free radicals via the cleavage of associations between the primary and secondary units of the condensed polyaromatic coal macromolecular network.

The SCCs obtained were found to have more volatile matter (VM) and easily degradable units than the OCs, resulting in increased tar formation from the SCCs. The various reactions that result in tar formation include depolymerization, dehydrogenation, decarboxylation, hydrogenation, and the stabilization of the heavyweight compounds obtained through secondary cracking reactions from the primary structure of coal. Thus, some information could be derived about the action of the e,N solvent system (NMP containing a small amount of EDA) studied through the Py-GCMS analysis of the samples. NMP, a polar diprotic solvent that has a good affinity for aromatics, is widely used in petroleum refineries. EDA has been found to be a good coal swelling and extracting solvent through the breaking of hydrogen bonds (Pande 2000; Pande and Sharma 2001, 2002). The synergistic action of the two solvents when used together, i.e., EDA, that cleaves the

H-bonds and NMP, extracts aromatics, and leaves SCCs with negligible ash content. Figure 4 summarizes the different compounds obtained from the OCs and SCCs. The highest amounts of mono-aromatics were obtained from the Bahula and Moonidih coals (both OC and SCC). NL OC gave mono-aromatics and alkenes predominantly, whereas NL SCC afforded relatively more n-alkanes and alkenes. The amounts of mono-, di- and tri-aromatics obtained during pyrolysis of the Bhagabandh SCC was relatively high as compared to the other coals (OCs and SCCs), indicating the effectiveness of the e,N solvent in extracting aromatics from this sample. Bhagabandh gave the highest extraction yield of all the coals. Based on these results, these coals and their SCCs could constitute a good source for the recovery of mono-aromatics such as benzene, toluene, xylene, and other substituted aromatic compounds.

When the area % of the five categories of the compounds were compared for their composition for coking coal (Moonidih coal) and non-coking high ash coal (Pandra coal), it was observed that the ash contents and the coking characteristics of coals affect the formation of certain coal macromolecules (Fig. 5a, b). The SCC (Fig. 5b obtained from the extraction of a coking coal Moonidih shows more monoaromatic content, whereas from the non-coking Pandra coal shows more alkenes. The tars of the SCCs obtained through the e,N solvent system show that the e,N extraction was largely aimed at the extraction of the specific moieties in coal; earlier studies have shown that the SCCs of the non-coking coals showed coking behavior extending their applicability in the steel industries (Pande 2000). Thus, the solvent extraction of the coals was found to enhance their coking properties by removing mono-aromatics and certain triaromatics, i.e., the use of mixed
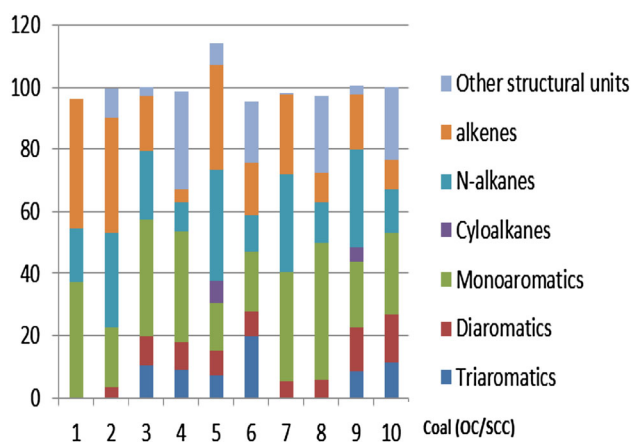
solvents may improve the coking properties of coal as well as significantly reducing the ash content.

3.2.2 PCA analysis

PCA was carried out for the OCs and SCCs of the five coals to identify structural similarities and differences between the samples before and after solvent extraction. Additionally, through PCA analysis, it was possible to detect zones showing variability in the data and detect any outliers/abnormalities (Melendez et al. 2012). PCA was performed with coal samples as samples (rows) and compound groups/compounds as features/variables (columns). Principal Component 1 (PC1) with an Eigen Vector corresponding to the largest Eigen Value of 354.35 captured 61.88% of the variance in the data. Similarly, Principal Component 2 (PC2) with an Eigen Vector corresponding to the second largest Eigen Value of 120.31 captures 21.01% of the variance in the data. Thus, the samples were represented on a 2D dimensional plane to visualize the spatial distribution relative to each other with axes PC1 and PC2 capturing a total of 82.89% of the variation in the data, when total peak area corresponding to compounds grouped as in Fig. 4, were considered as features. The loadings plot represents the weight coefficient of each compound group, denoting the contribution of original features (dimensions) to the directions of principal axes of variation in data (Principal Components/Eigen Vectors).

In Fig. 6b, *n*-alkanes and *n*-alkenes make the major contribution to PC1 in the positive direction, and the other compounds as a whole make significant contributions in the negative direction. Aromatic compounds have smaller weights than alkanes, alkenes, and other compounds, and they contribute to the negative side of PC1. Cycloalkanes lie close to the origin and have a small contribution in the positive PC1 direction. Aromatics have predominant contributions to PC2, with monoaromatic compounds having a high positive weight and diaromatics, triaromatics having negative weights. Alkenes and alkanes have smaller contributions in PC2 than aromatics, with nearly equal weights but in opposite directions, i.e., positive and negative PC2, respectively. The original coals cluster together on the positive side of the PC1 axis, whereas the super clean counterparts are spread on the negative side of PC1 except for Neyveli Lignite SCC, which shows significant similarity to the original coals in the direction of maximum variance in the data.

After removing Neyveli lignite from the analysis, in the scores plot, Fig. 7a, the SCCs of Bahula, Moonidih, Bhagabandh, and Pandra flip to the positive side of the PC1 in a pattern that is almost a mirror image about the PC2 axis. Since in the loadings plot, the n-alkanes, n-alkenes, monoaromatics, and others have flipped to opposite sides
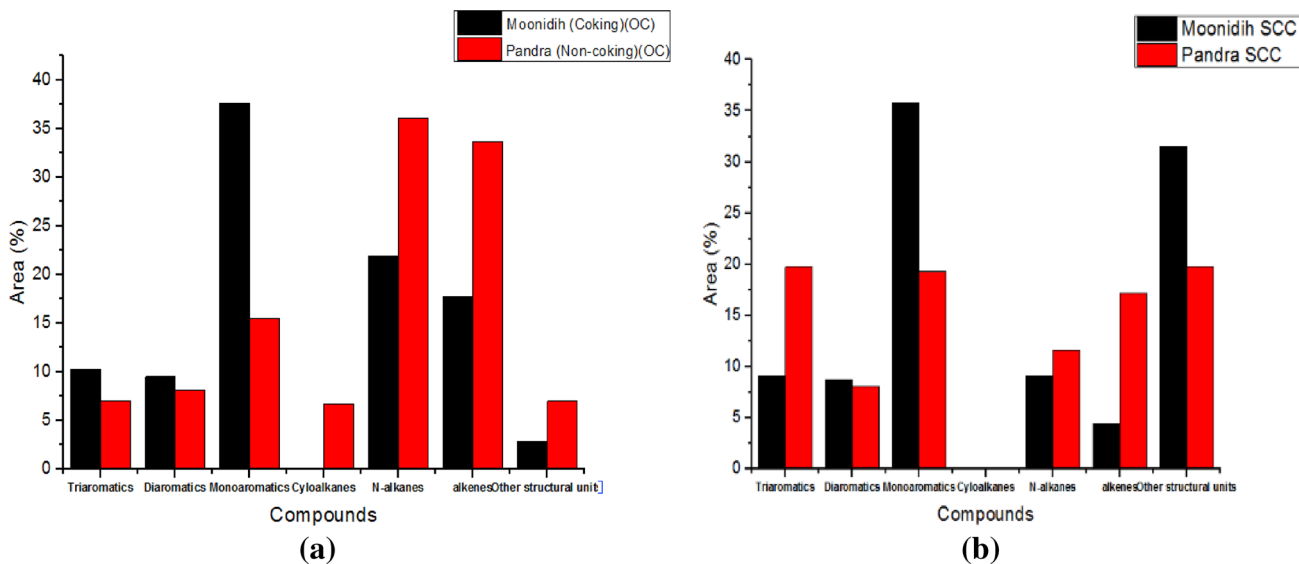

Fig. 4 Area (%) of the different compounds (y axis) in the tars of OCs and SCCs (1-NL OC, 2-NL SCC, 3-Moonidih OC, 4-Moonidih SCC, 5-Pandra OC, 6-Pandra SCC, 7-Bahula OC, 8-Bahula SCC, 9-Bhagabandh OC, 10- Bhagabandh SCC)

Fig. 5 Pyrolysis product comparison for **a** coking and non-coking OC **b** coking and non-coking SCC
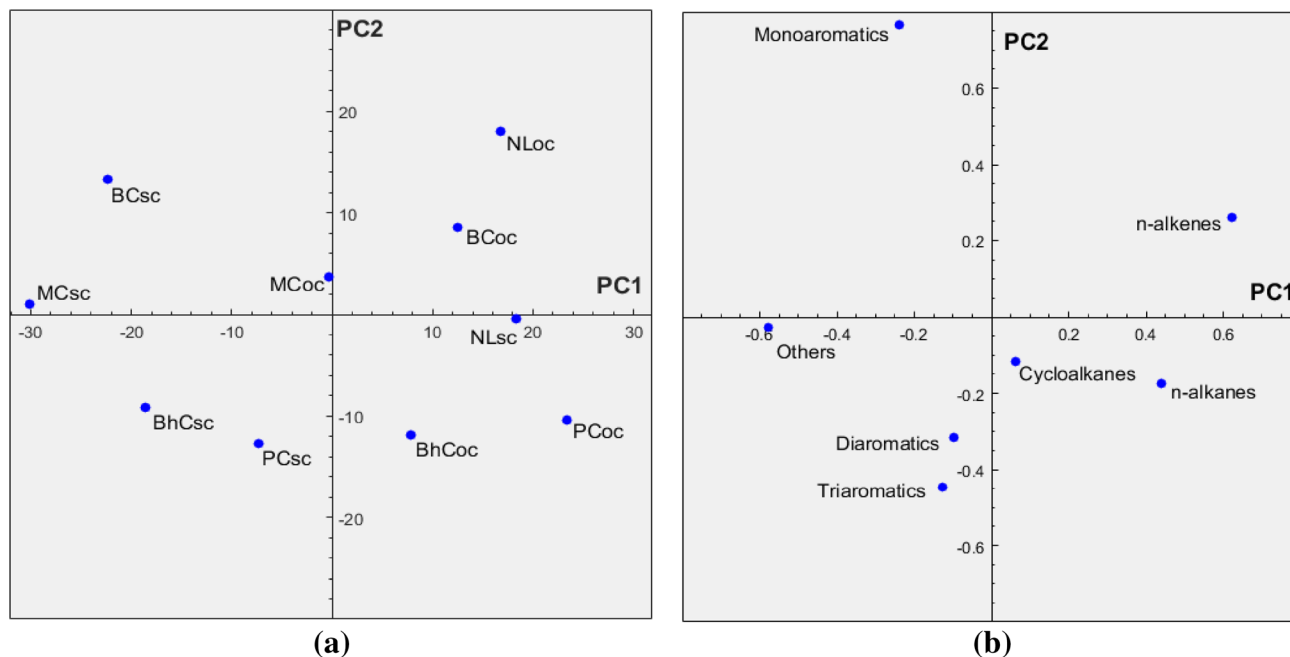

Fig. 6 Principal Component Scores Plot **a** and Loadings Plot **b** for Indian coals (OCs and SCCs)

too, the nature of the PC1 axis and its distinguishing characteristics based on the relative composition of samples is retained. The PC2 axis is still composed of the high positive weight of monoaromatics and negative weights of diaromatics and triaromatics. Diaromatics and triaromatics now have negligible loadings on PC1, and hence the specificity of PC2 increases and its nature is retained. Fig. 7 would be considered for further analysis.

PC1 distinguishes samples based on aliphatic and monoaromatics content, whereas PC2 primarily differentiates samples based on aromatic content. Original coals

have negative scores on PC1 and thus have high aliphatic content, comprising *n*-alkanes, *n*-alkenes, and cycloalkanes, as compared to solvent extracted super cleaned counterparts. For all samples, the major shift on PC1 as compared to a mild shift in PC2 scores supports the conclusions drawn by Sharma et al. (2019). The general trend of decrease in Aliphatic/Aromatic content scores post e,N solvent extraction for both non-coking high ash and coking low ash coals. Intense aromatics bands in SCCs signifying an increase in aromatics, as pointed out in ATR FTIR analysis of samples, can be observed from the fact that
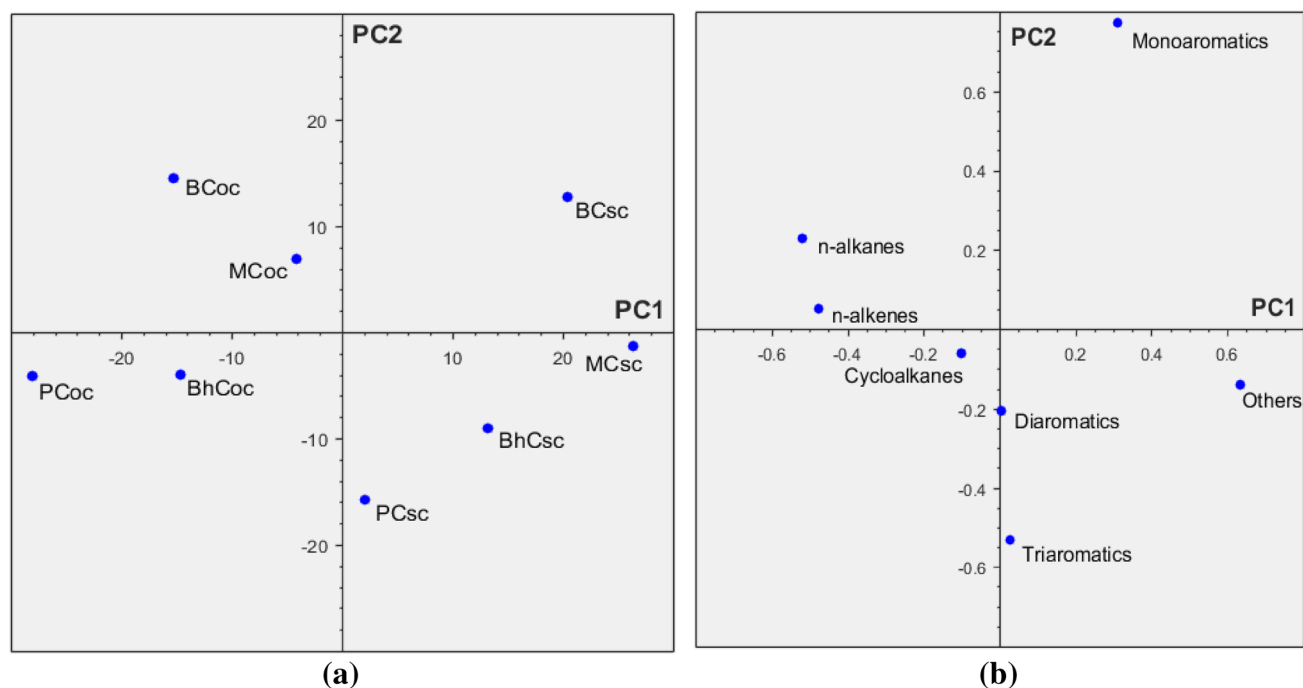
Fig. 7 Scores plot **a** and Loadings plot **b** excluding the Neyveli Lignite

monoaromatics have a positive loading on PC1 and coals' super cleaned samples having a higher negative PC2 score (shift towards more di- and triaromatics content) as compared to their original ones (Dhawan and Sharma 2019). This transformation was found to be more pronounced for non-coking coals, i.e., Bahula and Pandra Coals, for which average aromatic content increases and all other aromatic groups, as reported by Sharma et al. (2019). A significant increase in triaromatics content of Pandra Coal post solvent extraction was the highlight amongst these trends.

A total of 91 compounds can be identified through Py-GCMS of all samples together. The loadings and scores plot when PCA was done without clubbing compounds into groups is presented in Fig. 8. Compounds having loadings more than 0.1 on either PC1 or PC2 have been labeled, as their loadings are considered significant. Original and super cleaned coals form clear, distinct clusters on negative and positive halves of PC1, respectively. All compound plots were found to be consistent with the analysis drawn from grouped compound plots, although specific deductions are difficult to make, given the large number of weights represented in the loadings plot yet, in PC2, which comprises of significant loadings from both aromatics and aliphatic compounds, the distinguishing characteristics are not retained with both effects balancing each other making respective clean samples almost parallel to their raw ones. However, Bahula coals with a significantly high content of phenols have been clearly highlighted. For low ash coking coal Moonidih, having the highest content of polyaromatic

structures, total aromatics content has been reported to be compared before and after solvent extraction, whereas a distinct increase in other structural units like pyrrolidinones makes it super cleaned sample acquire high positive score on PC1.

3.2.3 Gaussian contour plots

As observed in Fig. 9, both contours are independent. The clusters are well separated, and thus, the shape of mixture contours does not differ much from the component contours, given that the weights (number of sample points) are comparable. Both SCCs and OCs are positively correlated (ellipsoidal shape tilted to the right), and their PC1 and PC2 scores co-vary in the given 2D plane. A positive correlation implies that for both OCs and SCCs, as the positive score increases in the PC1 axis, the positive score in PC2 increases. Moonidih, Bhagabandh, and Pandra OCs lie on the same contour curve and have the same probability density value. They belong to the original coals cluster under the same confidence limits.

4 Conclusions

Analytical pyrolysis results provide insights into the macromolecular structure of both coking and non-coking Indian original coals (OCs) and the selective action of the e,N solvent system, which produces products (SCCs) with
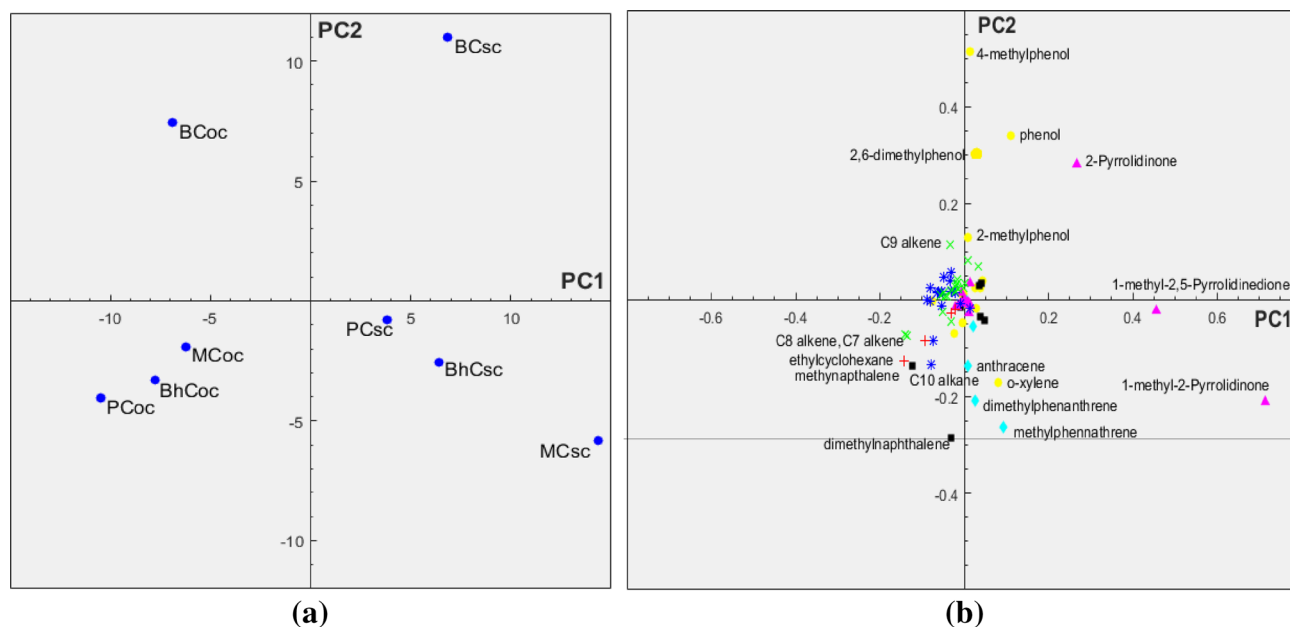
Fig. 8 PCA Scores Plot **a** and Loadings Plot **b** including all compounds (Blue Asterix—Alkanes; Red Plus-CycloAlkanes; Green Cross – Alkenes; Yellow Dot-Monoaromatics; Black Squares -Diaromatics; Cyan Diamonds- Triaromatics; Magenta Triangle-Others
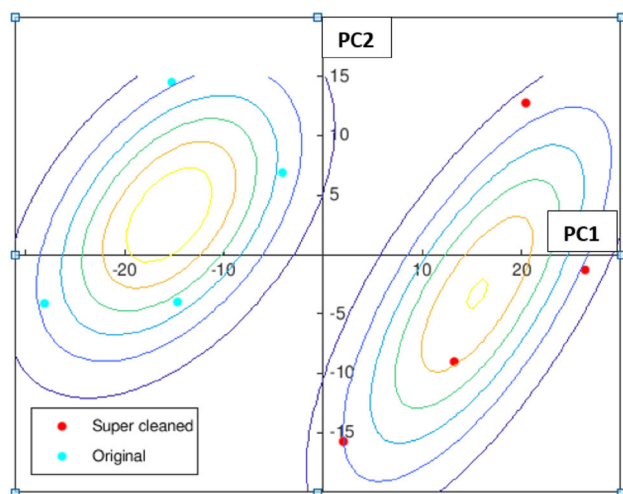

Fig. 9 Gaussian contour plot (excluding Neyveli lignite samples)

improved coking properties and greatly reduced ash content. These coals could act as good sources for monoaromatics production such as benzene, toluene, phenol, and substituted phenols, which have a wide array of industrial applications. Extensive Py-GCMS data analysis using PCA was performed on coal samples from five different Indian coal seams, namely, Pandra, Bahula, Moonidih, Bhagabandh, and Neyveli lignite and their corresponding SCCs. It primarily featured Scores and Loadings plots which expedited analysis of these samples based on the relative composition of constituent compounds through an accurate but compact visualization. This analysis of samples'

constituent compounds, initially, by categorizing based on various hydrocarbon classes, and later, based on individual compounds, was coherent and signified consistency of the mathematical formulation. The OCs and SCCs cluster on either side of the PC2 axis, except for NL-SCC, shows structural similarity to the OCs. Neyveli-Lignite has more volatile matter and moisture content and relatively lower ash content (5.2%) compared to the other coal samples, which may be attributed to its varied behavior from the other cluster of samples. Cluster characteristics and contour orientation were determined using the GMM algorithm. ATR-FTIR supported the inferences drawn from the Py-GCMS and PCA studies, showing the enhanced intensity of the aromatic bands in the SCCs as compared to OCs. Analysis of the ATR-FTIR spectra further led to the conclusion that the SCCs behave similarly for both coking and non-coking coals, verifying the selective action of the e,N solvent system on the polycondensed aromatic structures in the coals. Moonidih coal and its SCC were found to have the highest total content of aromatics.

Conflict of interest The authors declare no conflicts of interest.

References

Balachandran M (2014) Role of infrared spectroscopy in coal analysis—an investigation. Am J Anal Chem 5:367–372

Cooke NE, Fuller OM, Gaikwad RP (1986) FT-ir spectroscopic analysis of coals and coal extracts. Fuel 65:1254–1260

De Lira LFB, De Vasconcelos FVC, Pereira CF et al (2010) Prediction of properties of diesel/biodiesel blends by infrared spectroscopy and multivariate calibration. Fuel 89:405–409

Dhawan H, Sharma DK (2019) Refining of Indian coals to obtain super clean coals having insignificant amounts of deleterious elements under milder conditions. Miner Process Extr Metall, pp 1–12

Fardhyanti DS (2015) Analysis of coal tar compositions produced from sub-bituminous kalimantan coal tar. Int J Chem Mol Eng 9:1022–1025

Friesen WI, Ogunsola OI (1994) Principal component analysis of upgraded Western Canadian coals. Fuel Process Technol 38:139–151

Fujitsuka H, Ashida R, Miura K (2013) Upgrading and dewatering of low rank coals through solvent treatment at around 350 C and low temperature oxygen reactivity of the treated coals. Fuel 114:16–20

Harman-Ware AE, Crocker M, Pace RB et al (2015) Characterization of endocarp biomass and extracted lignin using pyrolysis and spectroscopic methods. BioEnergy Res 8:350–368

Huber GW, Iborra S, Corma A (2006) Synthesis of transportation fuels from biomass: chemistry, catalysts, and engineering. Chem Rev 106:4044–4098

Islas CA, Suelves I, Carter JF et al (2000) Pyrolysis-gas chromatography/mass spectrometry of a coal extract and its fractions separated by planar chromatography: correlation of structural features with molecular mass. Rapid Commun Mass Spectrom 14:1766–1782

Makan SR, Alexander JA, Solomon RA et al (2017) Physicochemical characterization of coal tar produced by pyrolysis of coal from Garin Maiganga and Shankodi deposits. J Miner Mater Charact Eng 5:288–297

Masayuki S, Masanori Y, Kensuke K, Kiyotaka T (2011) Reduction of GHG emissions by ash free coal utilization. Int J Sustain Energy Inst 1:22–30

Meier D, Fortmann I, Odermatt J, Faix O (2005) Discrimination of genetically modified poplar clones by analytical pyrolysis–gas chromatography and principal component analysis. J Anal Appl Pyrolysis 74:129–137

Melendez LV, Lache A, Orrego-Ruiz JA et al (2012) Prediction of the SARA analysis of Colombian crude oils using ATR–FTIR spectroscopy and chemometric methods. J Pet Sci Eng 90:56–60

Molina D, Uribe UN, Murgich J (2010) Correlations between SARA fractions and physicochemical properties with 1H NMR spectra of vacuum residues from Colombian crude oils. Fuel 89:185–192

Mullins OC (2010) The modified Yen model. Energy Fuels 24:2179–2207

Niedoba T (2014) Multi-parameter data visualization by means of principal component analysis (PCA) in qualitative evaluation of various coal types. Physicochem Probl Miner Process 50:575–589

Odeh AO (2015) Oualitative and quantitative ATR-FTIR analysis and its application to coal char of different ranks. J Fuel Chem Technol 43:129–137

Okuyama N, Komatsu N, Shigehisa T et al (2004) Hyper-coal process to produce the ash-free coal. Fuel Process Technol 85:947–967

Pande S (2000) Studies on Ethylenediamine assisted N-Methyl-2pyrrolidone extraction of coal scale up of the process and utilization of the super clean coal obtained

Pande S, Sharma DK (2001) Studies of kinetics of diffusion of N-methyl-2-pyrrolidone (NMP), ethylenediamine (EDA), and NMP+ EDA (1: 1, vol/vol) mixed solvent system in Chinakuri coal by solvent swelling techniques. Energy Fuels 15:1063–1068

Pande S, Sharma DK (2002) Ethylenediamine-assisted solvent extraction of coal in N-methyl-2-pyrrolidone: synergistic effect of ethylenediamine on extraction of coal in N-methyl-2-pyrrolidone. Energy Fuels 16:194–204

Pattiya A, Titiloye JO, Bridgwater AV (2010) Evaluation of catalytic pyrolysis of cassava rhizome by principal component analysis. Fuel 89:244–253

Rađenović A (2006) Pyrolysis of coal. Kem u Ind Časopis Kemičara i Kem Inženjera Hrvat 55:311–320

Schwarzinger C (2005) Identification of fungi with analytical pyrolysis and thermally assisted hydrolysis and methylation. J Anal Appl Pyrolysis 74:26–32

Sharma DK, Dhawan H (2018) Separative refining of coals through solvolytic extraction under milder conditions: a review. Ind Eng Chem Res 57:8361–8380

Sharma DK, Dhawan H, Morgan T, Crocker M (2019) Py-GCMS studies of Indian coals and their solvent extracted products. Fuel 256:115981

Sun M, Ma X-X, Yao Q-X et al (2011) GC-MS and TG-FTIR study of petroleum ether extract and residue from low temperature coal tar. Energy Fuels 25:1140–1145

Tesch S, Otto M (1995) Application of principal-component analysis to the interpretation of brown coal properties. Fuel 74:978–982

Van Krevelen DW (1993) Coal: typology, physics, chemistry, constitution. Elsevier, Amsterdam

Yan J, Bai Z, Bai J et al (2014) Effects of organic solvent treatment on the chemical structure and pyrolysis reactivity of brown coal. Fuel 128:39–45

Zhang D, Liu P, Lu X et al (2016) Upgrading of low rank coal by hydrothermal treatment: coal tar yield during pyrolysis. Fuel Process Technol 141:117–122