DE GRUYTER OPEN

***MINERALOGIA**, 48, No 1-4: 157-165 (2017)*
DOI: 10.1515/mipo-2017-0018
www.Mineralogia.pl

MINERALOGICAL SOCIETY OF POLAND
POLSKIE TOWARZYSTWO MINERALOGICZNE

Original paper

Data preprocessing and its influence on the adsorption capacity of calculated shale. Comparison of selected methods: A case study

Patrycja Waszczuk[1]*

[1] *Politechnika Śląska, Wydział Górnictwa i Geologii, Instytut Eksploatacji Złóż, Zakład Technologii Wydobywczych, Przeróbczych i Gospodarki Odpadami, ul. Akademicka 2, 44-100 Gliwice, Poland*
** Corresponding author*
e-mail: waszczukpatrycja@gmail.com

Received: April 21, 2017
Received in revised form: September 5, 2017
Accepted: September 5, 2017
Available online: September 30, 2017

Abstract. This article discusses the impact of choosing a data preprocessing method for the calculated gas density necessary to determine the sorption capacity of a material. The sample of gas-bearing shale was subjected to a volumetric sorption test. The obtained data, pressure and temperature were preprocessed by three methods: moving average, polynomial regression and locally weighted scatterplot smoothing. The results include the excess and absolute sorption calculated from data that were filtered, and data without pre-treatment and Langmuir isotherms' coefficients for every case.

Key-words: data preprocessing, sorption capacity

1. Introduction

Among other conditions for proper model building is properly collected, preprocessed and then processed data. Great emphasis in the daily routine of laboratories is placed on ensuring increasingly accurate sensors, data acquisition systems, devices and laboratory equipment providing the most optimum (usually the most stable) conditions of experiment possible. Of equal importance is the manner of dealing with the collected data. A properly chosen method of data preprocessing is crucial to obtain reliable, accurate and consequently valuable data.

Data obtained from sensors and collected by an acquisition system could be used directly in calculations, but their quality may (but not necessarily) be rather poor, thus not fully reflecting the essence of the phenomenon. Noise, outliers, and inconsistences can be found in raw sensor's data; therefore, that obtained data should be cleaned. Avoiding data preprocessing, the careless or ill-considered choice of data pre-treatment method can lead to inconsistencies in calculation, or even to the threat of serious error.

For the aforementioned reasons, data preprocessing should become a habit of every modern scientist, be an important part of every conducted experiment, and represent a special part of scientific projects, as well as every study connected with data acquisition.

The sorption capacity of shale is measured directly on a sample of material obtained from the considered shale bed, and then interpolated to the whole shale formation. The sample is often thermally treated to remove water and then crushed. This allows a gaseous phase contact with material on the largest area possible, while avoiding distortions in real sorption capacity caused by residual water contained in the pores of the tested material. Nevertheless, the crucial issue to be addressed is the problem of scale. The sample is nearly negligible compared with the volume of the shale formation. The smallest differences in sorption capacity extrapolated from the sample to the shale bed, transform into huge errors in the estimated amount of gas in the bed. This is a practical example for why data preprocessing in experiments investigating the sorption capacity in large-scale projects connected with geological structures (e.g. CO_2 capture and storage, methane extraction from gas-bearing shales or enhanced coalbed methane extraction) is so important.

2. Samples and methods

The Pomeranian gas-bearing shale was used in the conducted survey. Shale was obtained from the Early Paleozoic Baltic–Podlasie–Lublin Basin in Poland, considered as a source of natural gas and a site for CO_2 storage. The mineral composition of the tested sample is presented in Fig. 1 below, and only components which take part in sorption are detailed (TOC and CMC). The sample was powdered (< 0.1mm) and then dried at a temperature of 110°C (according to the norm PN-80/G-04511).

In the experiment, the manometric method of sorption measurement was used. A description of this method is often found in the literature (Kroos et al. 2002; Harpalani et al. 2006; Lutyński 2014). The data for calculation was collected from high-accuracy pressure and temperature transducers, connected to a computer and working in continuous mode online. Every step of the experiment was carried out over a 24 h period. The temperature and pressure were saved every second for 24 h, which helped to control whether the experiment was progressing properly. The setup is presented in Fig. 2.

The experiment comprised several stages. First, the tightness of the apparatus was checked by the injection of an inert gas (in this case, helium) under pressure of 180 bar for 24 h. If the gas tightness of the apparatus was acceptable (pressure drop not more than 0.02 bar per 24 h), the main part of the experiment started. After placing the sample in a vacuum (to allow helium to evacuate from the system), CH_4 was injected into the reference cell, and then left for a period of time (usually 1 h) necessary to stabilize its parameters (pressure and temperature). After stabilization, the valve between the sample cell and reference cell was opened and injection of gas into the sample cell proceeded.

After injection, the gas remained in the sample cell for 24 h. In the next step, the valve between the cells was closed and the reference cell filled with the next portion of gas. The stages of the experiment are presented in Fig. 3.

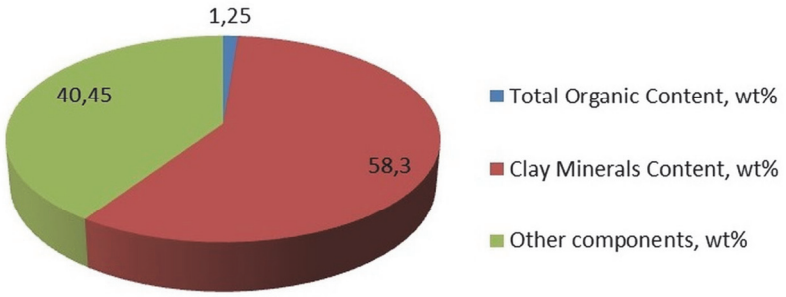

Fig. 1. Mineral composition of the investigated sample.

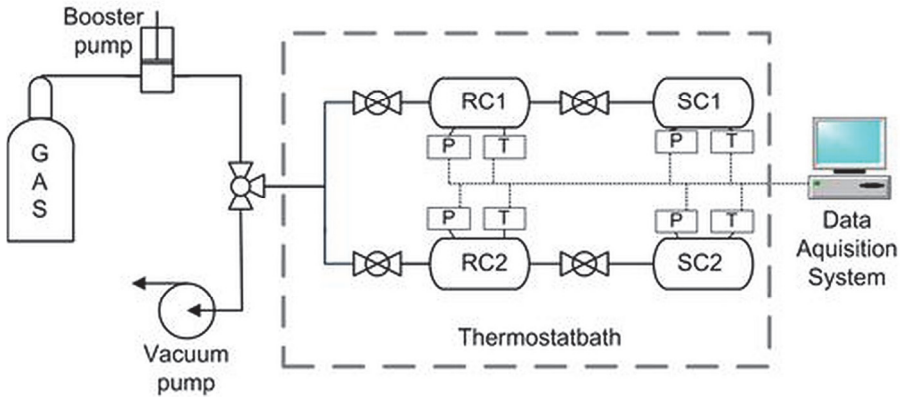

Fig. 2. Scheme of manometric setup: RC - reference cell, SC - sample cell, P - pressure transducer, T -temperature transducer (1 mbar or 0.05% FS).

To obtain reliable results, the equation of state for methane (Wagner, Span 1993) and helium (McCarthy, Arp 1990) was used. These equations are widely used and recommended in the literature (Van Hemert et al. 2009; Bush, Gensterblum 2011; Lutyński et al. 2011).

The data preprocessing was carried out in three ways: the moving average method, lowess (locally weighted scatterplot smoothing) and robust loess (local regression). These three methods have evolved from classical methods as a least square regression. Their advantage is that they are useful when a theoretical model of the process is not available. In the case of this study, the methods were selected arbitrarily due to their simplicity. The form of the obtained data also favored this method: data were large and dense, and there was much more data collected than necessary to perform a study; therefore, the problem of data loss in the moving average method could be neglected, and there was a negligible

number of outliers, which can influence the loess method but their impact was corrected with the robust loess method.

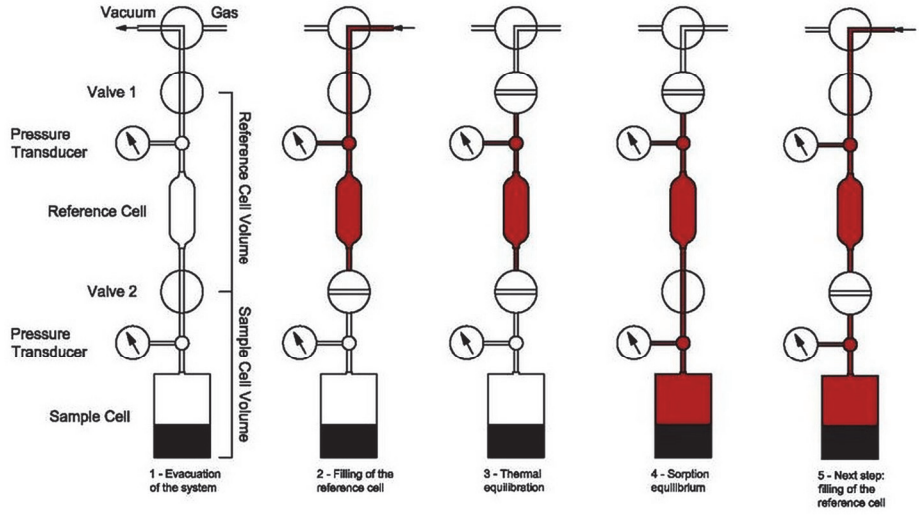

Fig. 3. Stages of the experiment.

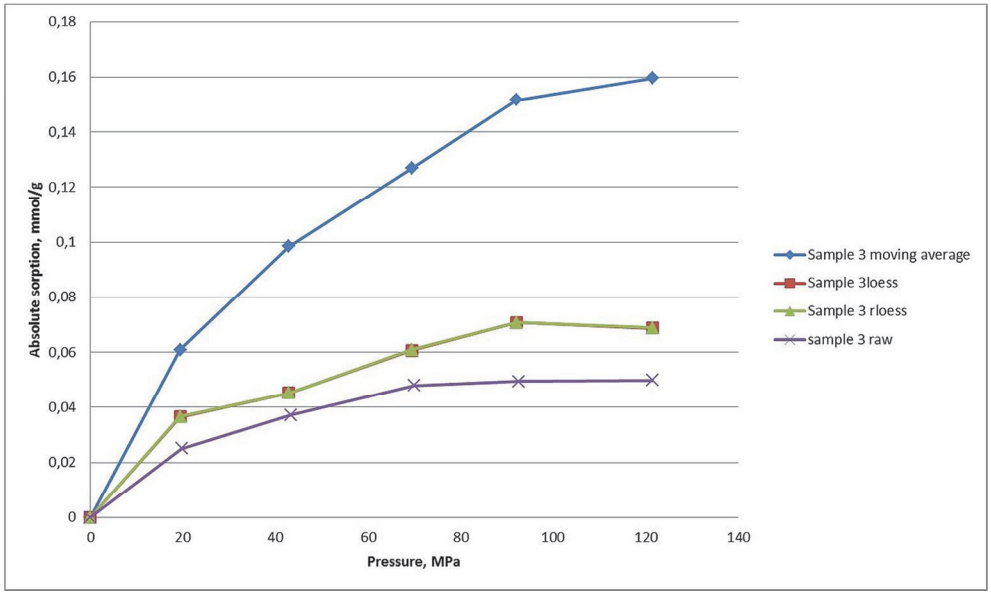

Fig. 4. Absolute sorption on sample calculated with raw and filtered data.

The data were processed in Matlab software with appropriate commands: smooth, loess and rloess. A smooth command was applied to the data moving average filter. The loess

command used local regression using weighted linear least squares and a 2nd degree polynomial model, while rloess was the same method as the local regression but assigned zero weight to data outside the six mean absolute deviations.

The processed data were then averaged and used to calculate the gas density. Gas density was calculated using the three-parameter Langmuir state (Sakurovs et al. 2007). Gas density is necessary to calculate the sorption capacity and predict the amount of gas borne by the shale. The equations utilized to calculate the capacity of shale are (Ambrose et. al. 2010):

$$G_{st} = G_a + G_f \tag{1},$$

where G_{st} is the total gas in-place, G_f describes the free gas in pores, and G_a describes the gas adsorbed on the surface of the micropores.

$$G_a = G_L \cdot \frac{p}{p+p_L} \tag{2},$$

where G_L is the Langmuir volume, p is the pressure and p_L is the Langmuir pressure.

$$G_f = \frac{32{,}0368}{B_g}\left[\frac{\emptyset(1-S_w)}{\rho_b} - \frac{1{,}318x\widehat{10^{-6}}M}{\rho_s}\left(G_L\frac{p}{p+p_L}\right)\right] \tag{3},$$

where Bg is the gas formation volume factor, ϕ is the porosity of the rock, Sw the water saturation, M the molar mass, ρb the bulk rock density and ρs the sorbed phase density.

The sorbed phase density is the part of equation which is crucial for this experiment, because its changes (and influence on the calculations) are under consideration.

3. Results

The results of the sorption when the data are applied to moving average, lowess and robust loess filters and raw data are presented in Fig. 5-8 and Table 1, 2 below:

TABLE 1

The pressure values after applying filters.

	Moving average (bar)	Loess (bar)	Rloess (bar)	Raw (bar)
Step 1	49.72	49.72	49.72	49.30
Step 2	77.96	76.96	76.96	77.92
Step 3	109.23	109.23	109.23	109.02
Step 4	126.73	126.73	126.74	126.54
Step 5	160.42	160.39	160.39	160.65

TABLE 2

Parameters of Langmuir model for CH_4 adsorption on sample.

Filter	V_L (mmol g^{-1})	ρ_s.(g per cm^3)	R^2
Moving average	0.6741	8.79	0.9974
Rloess/lowess	0.5350	8.24	0.9856
Raw	0.1801	17.14	0.9881

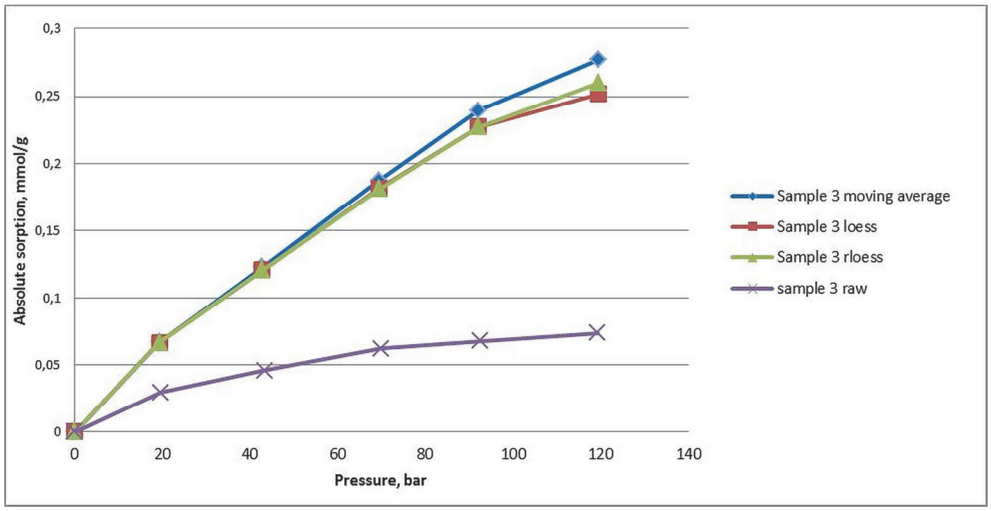

Fig. 5. Absolute sorption on sample calculated with raw and filtered data after applying patch.

The results were very surprising, and led to the question: Why are the results of the absolute sorption on filtered data so inflated? A small decrease was expected rather than any increase. Furthermore, the density of the sorbed phase returned implausible values, which implied a mistake. After taking a closer look at the programmed procedure, the problem was revealed: a filter was applying the whole matrix of data and then the whole matrix of results was averaged, which led to the incorrect results. This problem was solved in the following way: to check at what moment the sorption stops, the difference in consecutive results was calculated. If difference was less than 0.02 Mpa in the next 100 results, this meant that the sorption had finished. The influence of initial high pressure (when the sorption starts) was reduced. However, it was found that the calculated absolute sorption increased by approximately 2 times (Fig. 5).

The Fig. 6-8 present the excess and absolute sorption with the applied Langmuir model for the filtered and raw data.

At Fig. 7, it can be seen that the results are nearly the same for both filters and are unnoticeable on the chart (difference: 0.001 - 0.002 mmol g^{-1}).

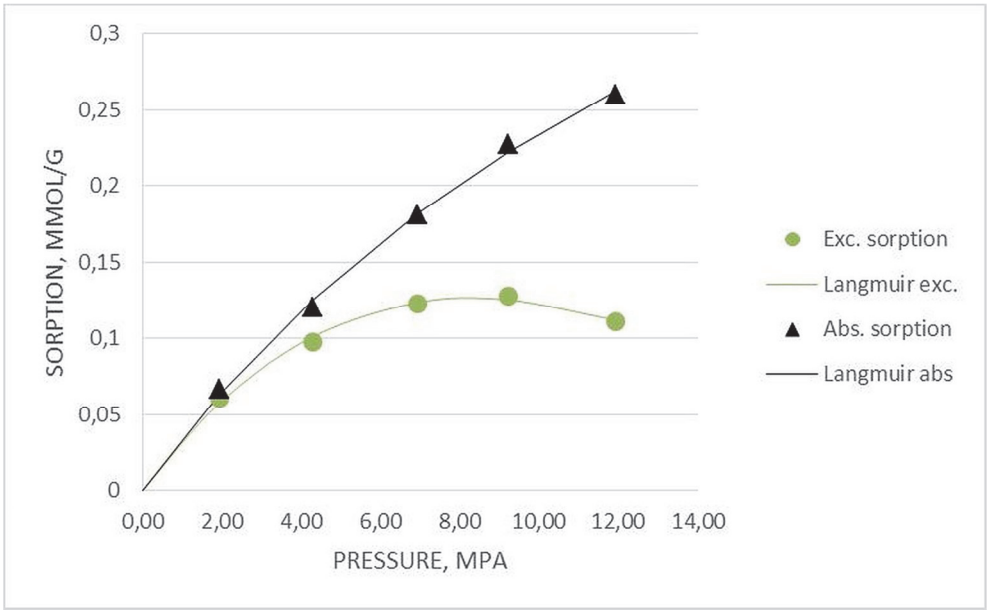

Fig. 6. Excess and absolute sorption with Langmuir model applied. Moving average filter used.

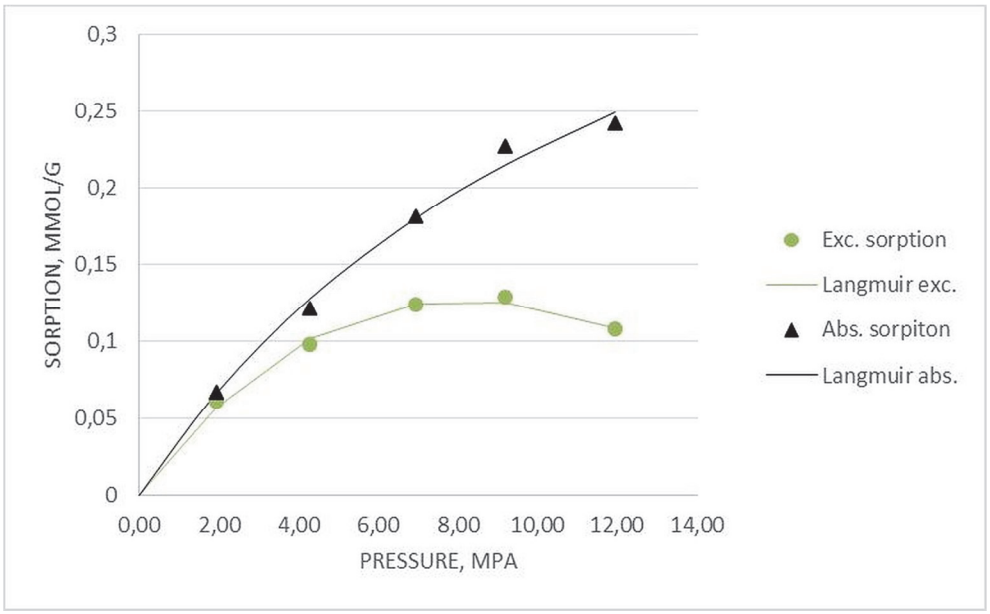

Fig. 7. Excess and absolute sorption with Langmuir model applied. Rloess/lowess filter used.

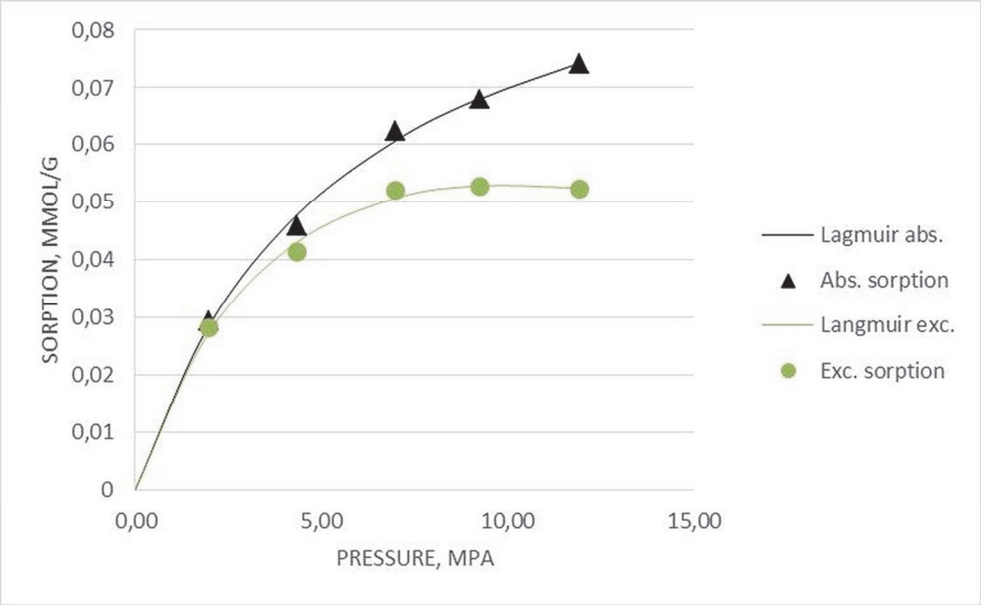

Fig. 8. Excess and absolute sorption with Langmuir model applied. Raw data used.

4. Discussion

Articles describing similar subjects in geology could not be found, and therefore discussion with related results is not possible here. Nevertheless, results of calculating gas density are available in the literature and are similar to those obtained by the author on raw data (Rexer et al. 2014). In this case, the most surprising result is that the calculated density (Table 1) using the raw data is nearly twice the size of that using the preprocessed data - at this point, more detailed studies are necessary.

The results of the excess sorption and absolute sorption are nearly the same in every method of preprocessing, but are higher than in the raw data (which was unexpected). The results are probably more precise due to the removal of noise and other inconsistences. Further investigation is necessary to determine if this is an error of calculation, because in the literature can be found values rather closer to those extracted from raw data than the preprocessed data (Khosrokhavar 2016).

The results obtained in this survey prove the exceptional role of data preprocessing in obtaining reliable data which reflect the existing situation. As mentioned in the Introduction, avoiding data preprocessing, the careless or ill-considered choice of data pre-treatment method can lead to inconsistencies in calculation, or even to the threat of serious mistakes, as shown in this case study. The obtained results deviated from the expected values, and were caused by the improper preparation of the code in the Matlab software. It can be clearly seen that paying insufficient attention to every step of the data processing can lead to flawed results. In fact, the mistake was not made at the stage of calculation, but at the programming stage.

5. Conclusion

Data preprocessing allows more accurate results to be obtained and helps to prevent mistakes, especially when it is necessary to interpolate results from the micro to macro scale. In this case, a difference in the final results is surprising, because the difference between the Langmuir volume is 3-times larger in the filtered data than the raw data, and the sorbed data density is twice lower in the filtered data than in the raw data. No reference bibliographic data could be found, and so the experiment remains to be verified. Paying insufficient attention to the choice of data processing may lead to incorrect results. Further work is necessary to determine if results in this study represent an error of calculation, or a new way to treat data in manometric method.

6. References

Ambrose, R. J., Hartman, R. C., Diaz-Campos, M., Yucel Akkutlu, I., & Sondergeld, C. H.. (New pore-scale considerations for shale gas in place calculations SPE Unconvencional Gas Conference, SPE 131722

Bush, A., & Gensterblum, U. (2011). CBM and CO_2-ECBM related sorption processes in coal. A review. *International Journal of Coal Geology, 87*(2), 49-71. DOI: 10.1016/j.coal.2011.04.011.

Harpalani, S., Prusty, B. K., & Dutta, P. (2006). Methane/CO_2 sorption modelling for coalbed methane production and CO_2 sequestration. *Energy & Fuels, 20,* 1591-1599. DOI: 10.1021/ef050434l.

Khosrokhavar, R. (2016). Mechanism for CO_2 sequestration in geological formations and enhanced gas recovery, Springer International Publishing Switzerland

Kroos, B. M., Van Bergen, F., Gensterblum, Y., Siemons, N, Pagnier, H. J. M., & David, P. (2002). High pressure methane and carbon dioxide adsorption on dry and moisture equiliobrated Pennsylvanian coals. *International Journal of Coal Geology, 74*, 69-92. DOI: 10.1016/S0166-5162(02)00078-2.

Lutyński, M. A., Battistutta, E., Bruining, H., & Wolf, K. A. A. (2011). Discrepancies in the assessment of CO_2 storage capacity and methane recovery from coal with selected equations of state Part I. Experimental isotherm calculation. *Physicochemical Problems of Mineral Processing*, *47,* 159-168.

Lutyński, M. A., & Gonzalez Gonzalez, M. A. (2016). Assessment of the total CO_2 storage capacity of shales - upscaling from sorption and reservoir data. 35th International Geological Congress 31ST august 2016, Cape Town, South Africa.

Lutyński, M. A. (2014). Impact of preparation and storage of activated carbon on the high pressure sorption of CO_2. *Buletin of Polish Academy of Sciences, 62,* 113-119.

Lutyński, M. A., & Gonzalez Gonzalez, M. A. (2016). Total gas in-place calculations for the Baltic-Podlasie-Lublin basin shales in Poland. E3S Web Conference. Mineral Engineering Conference MEC2016, Volume 8, 01053. DOI: 10.1051/e3sconf/20160801053.

Rexer, T., Mathia, E., Aplin, A. C., & Thomas K. M. (2014). High-pressure methane adsorption and characterization of pores in Posidonia Sales and isolated kerogens. *Energy & Fuels, 28,* 2886-2901. DOI: 10.1021/ef402466m.

Sakurovs, R., Day, S., Weir, S., & Duffy, G. (2007). Application of a Modified Dubinin−Radushkevich Equation to Adsorption of Gases by Coals under Supercritical Conditions. *Energy & Fuels, 21*(2), 992-997. DOI: 10.1021/ef0600614.

Van Hemert, P., Bruining, H, Rudolph, E. S. J., Wolf K, H. A., & Maas, J. G. (2009). Improved manometric setup for the accurae determination of supercritical carbon dioxide sorption. *Review of Scientific Instruments, 80*(3), 035103. DOI: 10.1063/1.3063064.