Multivariate Geostatistical Modeling of Lower Calorific Value in Multi-Seam Coal Deposits

Daphne Sideri 1,2, Christos Roumpos 1,2,*, Francis Pavloudakis 1, Nikolaos Paraskevis 1 and Konstantinos Modis 2

1 Department of Mining Engineering, Public Power Corporation of Greece, 10432 Athens, Greece; dafni.sideri@gmail.com (D.S.); f.pavloudakis@dei.com.gr (F.P.); nikolasparaskevis80@gmail.com (N.P.)
2 School of Mining and Metallurgical Engineering, National Technical University of Athens, 15780 Athens, Greece; kmodis@mail.ntua.gr
* Correspondence: c.roumpos@dei.com.gr; Tel.: +30-69-7979-9291

Received: 6 August 2020; Accepted: 4 September 2020; Published: 7 September 2020

Featured Application: This research is significant for the mining industry since a reduction to estimation uncertainty at no additional drilling cost could lead to critical savings in coal production.

Abstract: The estimation of fuel characteristics and spatial variability in multi-seam coal deposits is of great significance for the optimal mine planning and exploitation, as well as for the optimization of the corresponding power plants operation. It is mainly based on the quality properties of the coal (i.e., Lower Calorific Value (LCV), ash content, CO2, and moisture). Even though critical, these properties are not always measured in practice for all available borehole samples, or, they are generally estimated by using non-parametric statistics. Therefore, spatial modeling of LCV can become problematic due to the limited number of data. Thus, the use of other available correlated attributes might be helpful. In this research, techniques of multivariate geostatistics were used to estimate and evaluate the spatial distribution of quality properties in a multi-seam coal deposit, with special reference to the LCV. More specifically, kriging, cokriging, and Principal Component Analysis (PCA) techniques were tested in a case study as estimators of the LCV, using an extensive set of borehole data from the South Field lignite mine in Ptolemais, Greece. The research outcomes show that the application of kriging with two PCA factors and the use of inverse transform result in the best LCV estimates. Moreover, cokriging with two auxiliary variables gives more accurate values for a LCV estimate, in relation to the kriging technique. The research outcomes could be considered significant for the coal mining industry, since the use of correlated quality attributes for the estimation of LCV may contribute to a reduction of the estimation uncertainty at no additional drilling cost.

Keywords: geostatistical modeling; ordinary kriging; cokriging; principal component analysis; multi-seam coal deposits; quality parameters

1. Introduction

Greek lignite deposits have a multiple-layered structure consisting of several coal seams which are separated mainly by calcareous and argillaceous waste beds (Figure 1). The necessity of selective mining due of the above stratigraphic deposit structure, combined with the requirements for high production rates, was the reason for the application of the continuous surface mining method for more than sixty years [1].

The modeling and evaluation of multi-seam coal deposits and, consequently, the estimation of the mineable coal reserves and their quality properties is a process of compositing seams into blocks of exploitable coal seams and separating them from the blocks of waste (non-lignite) material [2]. This
process combines with spatial block modeling, as well as with geologic modeling in order to estimate the spatial variability of the coal reserves and the corresponding energy content. The main criteria used for the formation of the exploitable compositing coal blocks include the minimum thickness of coal blocks and waste layers that can be selectively excavated by the mining equipment, the cut-off quality properties of the run of mine coal (e.g., maximum ash content and minimum Lower Calorific Value (LCV)) and the dilution with waste and mining lignite loss at the top and bottom of each block. The composite blocks of exploitable coal may include thin waste layers, while the blocks of waste material may include thin coal seams which cannot be selectively excavated by the available equipment. The composition of seams into blocks of exploitable coal and waste material is a computer-aided iterative process based on suitable algorithms [3,4].

The analysis of the compositing process indicates that the accurate determination of the exploitable coal blocks, based on the original drill-hole data, and the spatial modeling of a coal deposit have a significant impact on the successful strategic planning of the corresponding mining project [5]. In continuous surface mining projects for the exploitation of multi-seam coal deposits, the investigation of the parameters that affect the coal deposit modeling and the evaluation results are significant, considering the complicated and continuously varying mining conditions and the requirements for highly efficient exploitation technologies and enhanced recovery of the mineral resources [2]. In such projects, there is an inherent uncertainty mainly related to the spatial variability of the mineral deposit parameters. This uncertainty should be taken into account in all stages of mine planning and design as well as in production scheduling and the economic evaluation of the projects [6]. The main challenges refer to the deposit modeling in a way that allows a quantitative understanding of the spatial variability of the fuel characteristics and the related uncertainties, which in turn helps to optimize mine exploitation and to reduce fluctuations in the quality of the run-of-mine ore [7].

These aspects of multi-seam coal deposits evaluation clearly indicate that the quality parameters of coal have a significant effect on the adoption of the criteria for the calculation of the exploitable blocks. However, parameters such as the LCV of the coal deposits often are not measured for all borehole

Figure 1. Multiple-layered structure of Greek lignite deposits.
samples taken during the exploration stage [8]. Therefore, their estimation must be primarily based on careful modeling by means of deterministic or probabilistic approaches [9–11].

Generally, natural heterogeneity in relation with measurement errors introduces extensive sources of uncertainty in coal characteristics modeling. In this line, various research studies have indicated the uncertainty which is related to the natural heterogeneity of coal [4,12]. In these studies, non-parametric statistics are generally used to correlate different coal properties. However, in more recent studies, geostatistical techniques are applied for the spatial estimation of coal parameters [1,7,8,13,14]. The kriging method is the most popular for modeling the spatial variations of coal quality characteristics all over the world [15–17]. The method has been modified in many ways in order to produce more accurate results, such as cokriging and factorial cokriging [18], kriging with external drift or auxiliary variable [14], or covariance matching constrained kriging [19]. There are also a few articles that report the use of geophysical measurements as auxiliary data in order to improve coal estimation accuracy [14,20–22].

In this line, the objective of the present study is to improve the estimation of the spatial variability of LCV by using other quality attributes as auxiliary variables. More specifically, multivariate geostatistical techniques (i.e., kriging, cokriging, and Principal Component Analysis (PCA)) are coupled to evaluate the quality parameters of coal, with particular reference to the LCV and with the support of other correlated quality attributes. The use of the correlated quality attributes for the estimation of LCV could be considered significant for the mining industry, since a reduction to estimation uncertainty could be achieved at no additional drilling cost.

2. Methods

The Lower Calorific Value is considered as an important quality parameter for the estimation of coal reserves. However, the LCV of a coal deposit often is not measured for all available borehole samples and should thus be estimated from other measured parameters.

The aim of the geostatistical analysis performed in this research is to determine the best estimation conditions of the Lower Calorific Value, taking into account all the available information concerning the quality parameters of coal, provided in the data set.

The main issues faced in our analysis are the following:

- The examination of the LCV correlation to the other available coal properties in order to determine the auxiliary variables which would be used for the estimation.
- The utilization of proper methodologies to ease the inclusion of these auxiliary variables in the estimation procedure. Among such methodologies, kriging and cokriging are geostatistical tools that exploit the spatial structure in order to map the objective variable on a given domain [18,23]; kriging by taking into account measurements of the objective variable in the neighborhood of the estimation points, while cokriging by considering also auxiliary variables [23,24]. However, as long as cokriging with multiple variables may not in fact be feasible, a third method was examined, where quality parameters were replaced with their principal factors derived from Principal Component Analysis [25]. Only 8% of the total variability were missing as a result of this substitution (see following section), thanks to the overabundance of information.
- The adoption of the most convenient method to employ, according to its performance on the present case study. This was achieved by cross-validating the results from all the examined methods [26]; first by estimating the points where LCV is already known and then by comparing the results to the measured values according to appropriate statistics.

Before proceeding further, the above methodologies are briefly outlined in the following paragraphs:

2.1. Principal Component Analysis

Principal Component Analysis (PCA) is a methodology for analyzing data in higher dimensions where graphical representation is not an option. The main idea is to compress the data by reducing their dimensionality without losing much information. In practice, an array of usually correlated
variables is linearly transformed into an equal number of independent variables, with standard normal
distribution, called factors [25]. Since the transformation is linear, the initial variables can be generally
described as linear combinations of these factors. The important thing is that due to overabundance of
information owing to possible intercorrelations between variables, only a few factors are generally
enough to explain most of the original variability.

The aim of local estimation is to calculate the most probable value of a random variable in a
location inside a homogeneous zone of the study area [23]. The information used for the estimation is
generally comprised of a set of samples (e.g., n core measurements) and a structural knowledge (e.g., a
covariance model accounting for the spatial variability in the modeled area) [23].

2.2. Ordinary Kriging

Ordinary kriging [23,24] is a procedure for local estimation which calculates the best linear
unbiased estimator of the modeled attribute. The restriction to the category of linear estimators means
that only second-order moments of the random function (i.e., the covariance function or the variogram)
are considered. This is because it is easier to derive these moments from the available data in practice.

In practice, to estimate the value of an attribute \(Z_0\) in a point \(x_0\) of the space using a set of
measurements on \(n\) nearby locations \(x_\alpha\) (\(\alpha = 1, n\)), it is necessary to form a linear combination of
the type:

\[
Z_0^*(x_0) = \sum_{\alpha=1}^{n} \lambda_\alpha^0 Z_0(x_\alpha),
\]

(1)

where the matrix of coefficients \(\{\lambda_\alpha^0\}\) is calculated so as to ensure the unbiasedness of the estimator and
the minimization of the difference between the unknown and the estimated (kriged) values on each
estimation point. Minimization accounts also for the spatial variability of the attribute, as expressed by
the variogram function \(\gamma_0(h)\), which equals the semivariance of the variable increment by the distance \(h\):

\[
2\gamma_0(h) = \text{var}[Z_0(x + h) - Z_0(x)].
\]

(2)

A theoretical variogram model is then adapted to the calculated experimental variograms in order
to be used in the kriging algorithm. Another important characteristic of kriging is that it provides an
estimation error variance, which can be used as a measure of its accuracy. Furthermore, estimation
error variance can also be used to identify areas requiring supplementary sampling.

2.3. Cokriging

Cokriging [23,24] on the other hand is used when the target variable is under-sampled so as it
is not possible to ensure acceptably accurate estimates. An improvement to the estimation accuracy
may then be provided by considering the spatial correlations of this variable to other, more adequately
sampled attributes. In order to estimate the value of an attribute \(Z_0\) in a point \(x_0\) of space using a set
of values measured on \(n_0\) nearby points and additionally a different set of measurements of other
auxiliary parameters \(Z_i\) (\(i = 1, p\)) spatially correlated to \(Z_0\), a linear combination of all these data is used
again. If \(n_i\) (\(i = 1, p\)) is the number of sampled points near \(x_0\) and if the \(p + 1\) attributes are considered,
the cokriging estimator is

\[
Z_0^*(x_0) = \sum_{i=0}^{p} \sum_{\alpha=1}^{n_i} \lambda_{i\alpha} Z_i(x_\alpha).
\]

(3)

A minimization analogous to ordinary kriging is then employed to determine the coefficient
matrix \(\{\lambda_{i\alpha}\}\), which characterizes the spatial variability of all attributes by defining the set of direct and
cross variograms \(\gamma_{ij}(h)\). The generic cross variogram between attributes \(Z_i\) and \(Z_j\) equals again the
semicovariance of the pair increments by distance \(h\):

\[
2\gamma_{ij}(h) = \text{cov}[(Z_i(x + h) - Z_i(x)), (Z_j(x + h) - Z_j(x))].
\]

(4)
Variograms are brought into the estimations by means of a theoretical function that ensures mathematical consistency. The most commonly used model is the linear model of coregionalization, where all direct and cross variograms are deduced from linear combinations of \( n \) elementary direct variograms \( \{\gamma_k(h), k = 1 \text{ to } n\} \), i.e.,

\[
\gamma_{ij}(h) = \sum_{k=1}^{n} b_{ij}^k \gamma_k(h) \quad \text{with} \quad b_{ij}^k = b_{ji}^k, \quad \forall k,
\]

(5)

where \( \forall k \) the matrix of coefficients \( \{b_{ij}^k\} \) is conditionally non-negative definite.

3. Geology of the Study Area

Greece was considered as a large producer of soft brown coal, producing about 45–70 Mt of lignite per annum, mainly at the Lignite Center of Western Macedonia. The annual lignite production during the period 2016–2018 ranged to 35 Mt, while it was reduced to 26 Mt in the year 2019. Since 2020, the lignite production progressively decreases, as a result of the gradual mine closure.

The Lignite Center of Western Macedonia operates five mines: Mavropigi, Kardia, and the South Field in the Ptolemais area, and Amyndeon and Lakkia in the Amyndeon area. The coal deposits in the above areas have been estimated approximately at 660 Mt. In Figure 2, an overview of the mines in the Ptolemais area is shown.

![Figure 2. Mine boundaries in the Ptolemais lignite basin (June 2019).](image)

The South Field mine (SFM) is the largest of the five operational surface mines at the Lignite Center of Western Macedonia, with an overall annual production of about 7–15 Mt, feeding mainly the Agios Dimitrios Power Plant. Mining operations in the South Field mine started in August 1979. For operational purposes, the SFM has been divided into 11 sectors and currently operates on eight benches in Sector 6 and four in Sector 7, mainly using the continuous mining method, which uses bucket wheel excavators (BWEs), conveyors, and stackers.
As it is mentioned in the research of Roumpos et al. [1], the deposit of the SFM is divided into partly narrow faulted blocks by several faults, which mainly strike in a NW-SE to WNW-ESE direction and mostly consist of down thrown and, to a lesser extent, of up thrown faults with slight faulting in the underlying seams. The overall thickness of the lignite-bearing series amounts to approximately 20–30 m in the southeast and northeast of the mining field, while towards the west and north-west, it increases to approximately 70–100 m in the present opencast mine slope. In the areas of the western and eastern final slopes, the thicknesses of the split seams partly decrease over a short distance, with the intercalations increasing in thickness at the same time.

The overlying strata thickness ranges from approximately 70–100 m at the present mine rim to up to approximately 200 m in the southern part of the mining field. The layer series is composed of finely clastic, sandy clays and calcareous, clayey marls with coarsely clastic, marly sands, sandy gravels, and pebbles (Figures 3 and 4i,ii).

![Figure 3. Geological map of the South Field mine.](image-url)
Figure 4. Geological setting of the South Field mine, (i) typical geological section AA’ (the trace of cross-section AA’ is presented in Figure 3), (ii) stratigraphic column (modified from Roumpos et al. [1]).

Water outlets from the coarsely clastic sediments have been observed in the opencast mine. Furthermore, these layers locally incorporate consolidated intercalations seriously impeding mining operations, which principally consist of conglomerates, sandstone, and calcareous marl [27].
4. Dataset

Throughout the South Field mine, several hundreds of sampling boreholes were drilled by the Public Power Corporation of Greece—PPC (Figure 5). During the present study and based on the above, all the borehole data were analyzed and interpreted in order to create a dataset including the characterization of distinguished horizons and their range of main quality properties. This dataset includes information on 66,376 samples taken from 825 boreholes. Sample lengths range from 0.01 to 19.43 m, with an average value of 1.13 m. Only 143 samples are smaller than 0.1 m and 48 samples are greater than 4 m.

The selected samples have a set of information comprising the four most relevant quality properties—moisture percentage (%), ash percentage (%), CO\textsubscript{2} content (%), and Lower Calorific Value (kcal/kg). However, not all attributes are measured on all samples. More specifically, LCV, which is the main variable of interest, has been measured only in a portion of the total number of samples, as shown in Table 1. Usually, samples with LCV values are also analyzed for moisture, ash, and CO\textsubscript{2}.

<table>
<thead>
<tr>
<th>Quality Properties</th>
<th>Number of Samples</th>
<th>Minimum</th>
<th>Maximum</th>
<th>Mean</th>
<th>Std. Dev</th>
</tr>
</thead>
<tbody>
<tr>
<td>Moisture (%)</td>
<td>23,027</td>
<td>17.00</td>
<td>96.80</td>
<td>54.62</td>
<td>4.65</td>
</tr>
<tr>
<td>Ash percentage (%)</td>
<td>23,027</td>
<td>4.71</td>
<td>81.60</td>
<td>28.35</td>
<td>10.40</td>
</tr>
<tr>
<td>CO\textsubscript{2} content (%)</td>
<td>9,136</td>
<td>0.10</td>
<td>71.40</td>
<td>16.99</td>
<td>9.66</td>
</tr>
<tr>
<td>Lower Calorific Value (kcal/kg)</td>
<td>11,186</td>
<td>93.00</td>
<td>2610.00</td>
<td>1419.59</td>
<td>306.41</td>
</tr>
</tbody>
</table>

The large number of analyzed samples per borehole is due to the multi-seam structure of the lignite deposit and the significant variation of the main quality characteristics between these seams (Figure 6). Nevertheless, the lignite seams need to be excavated either separately or together, depending on the thickness of the waste seams that are intercalated between them. The co-excavation of lignite and waste seams of limited thickness is considered as the optimum operating regime for increasing the
productivity of the high capacity bucket wheel excavators in spite of the deterioration of the quality characteristics of the mined lignite.

Figure 6. Variation of Lower Calorific Value (LCV) of lignite seams separated by waste seams (clay, sand, and marl), (a) typical multi-seam bench of the South Field mine (AB: part of the drill-hole 122/304), (b) analytical data of the drill-hole 122/304, (c) waste and coal seams of the drill-hole 122/304 (location of the drill-hole is presented in Figure 3).
5. Results and Discussion.

5.1. Structural Analysis and Spatial Modeling

Geostatistical models work in general by considering the probabilistic nature of the underlined variable but also account for the spatial dependency of observations. According to the above assertion, these algorithms adopt a certain spatial correlation of the natural variables, which, in all other ways, are considered as random processes. This assumption allows a certain degree of interdependence between model values at different spatial locations [28].

As stated above, the spatial inhomogeneity and local absence of information in the dataset prompts to the adoption of auxiliary variables for mapping LCV values. The choice of the proper quality characteristics to be employed as auxiliary variables is done by considering the correlation of LCV to the other available attributes. The results of this test indicate a significant correlation between ash, CO$_2$, and LCV. On the other hand, moisture is poorly correlated to LCV (Figure 7).

![Figure 7. Scatter plots of LCV versus CO$_2$, moisture, and ash.](image)

In order to study the spatial dependencies of LCV values, the experimental variogram was calculated with a lag size of 1 m on the vertical direction and 330 m on the horizontal plane (Figure 8). A model with two spherical structures was adjusted to the data.

Moreover, to investigate the cross-correlation between the quality parameters in space, it was necessary to study the individual variograms and the cross variograms of the above variables (Figures 9 and 10). A coregionalization model consisting of two spherical structures was adjusted on the experimental variograms, which were calculated with a lag size of 10 m on the vertical direction and 250 m on the horizontal plane. Sill matrixes were adjusted through conditional fitting of the experimental variograms. Experimental variograms show a range of about 20 m in vertical and 1300 m in horizontal direction.
Figure 8. Experimental and fitted anisotropic variogram models for LCV. Dashed polylines (blue color) represent experimental variograms; thin dashed lines (black color) represent variance; continuous lines (red color) represent variogram models.

Figure 9. Experimental and model cross variograms for LCV, ash, and CO$_2$ in the horizontal directions. Experimental structure is modeled by an isotropic spherical variogram. Dashed polygonal lines (blue color) symbolize experimental variograms; thin dashed lines (black color) symbolize variance and covariance; continuous lines (red color) symbolize variogram models.
Figure 10. Experimental and model cross variograms for LCV, ash, and CO₂ in the vertical direction. Experimental structure is modeled by an isotropic spherical variogram. Dashed polygonal lines (blue color) symbolize experimental variograms; thin dashed lines (black color) symbolize variance and covariance; continuous lines (red color) symbolize variogram models.

Principal Component Analysis was applied to reduce the dimensionality of the dataset and to synthesize the necessary information by using independent components. This way, cokriging could be avoided. As seen in Figure 11, the significance of each factor, as revealed by the weight of its eigenvalue, is expressed by its percentage of the total variance. Due to the overabundance of information due to inter-correlations among the quality parameters, the first two PCA factors explain almost 92% of the total variance.

Figure 11. Projection of the quality parameters on the plane spanned by the first two factors of the Principal Component Analysis (PCA).
The variogram models for Factors 1 and 2 (Figures 12 and 13) were fitted as double spherical structures with a lag size of 1 m on the vertical direction and 330 m on the horizontal plane. Variogram model ranges are 700 and 1400 m in the horizontal plane and 25 m in the vertical direction.

**Figure 12.** Experimental and fitted anisotropic variogram models for Factor 1. Dashed polylines (blue color) represent experimental variograms; thin dashed lines (black color) represent variance; continuous lines (red color) represent variogram models.

**Figure 13.** Experimental and fitted anisotropic variogram models for Factor 2. Dashed polylines (blue color) represent experimental variograms; thin dashed lines (black color) represent variance; continuous lines (red color) represent variogram models.

Interpolation with kriging Equations (1) and (2) and cokriging Equations (3)–(5) was performed in a 100 m × 100 m × 10 m horizontal orthogonal parallelepiped grid, divided into 100,000 block elements. The searching neighborhood used was of ellipsoid type, with a maximum distance of 1500 m in the horizontal directions and 10 m in the vertical direction. Cokriging was applied by considering LCV as target attribute with ash and CO₂ as auxiliary variables.

The interpolation results are shown in Figures 14 and 15.
In order to model LCV with PCA, the first two factors produced by the transformation of the original dataset were kriged on the same grid as in the previous models. After this step, an LCV value was assigned on all grid elements by the inverse transformation of the kriged factors. As seen in Figure 16, the model area in this case is slightly smaller than the one generated by the previous methods. This is because PCA applies only on samples where the full set of attributes is measured.
5.2. Cross-Validation of the Estimation Methods

In order to validate the three interpolation methods for LCV modeling, cross-validation results were compared by employing the correlation coefficients between estimated and measured values. Cross-validation results, which are reported in Figures 17 and 18, can be summarized in the following points:

1) Cokriging, which accounts for both spatial correlation and two auxiliary variables, produces a more accurate LCV model than kriging;

2) PCA, which takes into account all the available attributes, showed the highest correlation between true and estimated values of LCV.

![Image of 3D model of LCV (kcal/kg) spatial distribution produced by the application of PCA.](image)

**Figure 16.** 3D model of LCV (kcal/kg) spatial distribution produced by the application of PCA.

![Scatter plot of measured vs. estimated Lower Calorific Value (LCV) with (a) kriging and (b) cokriging.](image)

**Figure 17.** Scatter plot of measured vs. estimated Lower Calorific Value (LCV) with (a) kriging and (b) cokriging.
Moreover, in order to further compare the performance of the previous methods, we plot the statistical distributions of the LCV from samples and the LCV estimated with these methods. The results are shown in Figure 19 where it is apparent that the estimation with PCA is closer to reality.

**Figure 18.** Scatter plot of measured vs. estimated LCV with PCA.

**Figure 19.** Statistical distributions of LCV from (a) samples, (b) kriging, (c) cokriging, and (d) PCA.
6. Conclusions and Future Research

The accurate estimation and representation of quality properties is a key factor for the successful calculation of mineable coal reserves in multi-seam deposits. The usual approach for assessing the spatial distribution of these quality properties is by the use of non-statistical interpolation methods. In this research, the study of the quality properties of lignite in the South Field mine, Greece, has shown that it is possible to estimate LCV values by using CO\textsubscript{2} and ash content as auxiliary variables. Cokriging, which accounts for the spatial structure of the underlying phenomenon and also for some of the additional attributes, provided a better estimation of LCV compared to kriging, as shown by cross-validation of the results. Furthermore, PCA gives even better results because it takes into account the full set of the available attributes. The above results are significant for the mining industry, since a reduction to estimation uncertainty at no additional drilling cost could lead to critical savings in the cost of coal production.

In consideration of further improving estimation results, especially when accounting for distant model locations, the additional use of related attributes like geological properties or relative positions and proportions of lithofacies might be of prominence. Therefore, a future step of the research would be to consider the above parameters through the adoption of current geostatistical simulation methods (e.g., Plurigaussian simulation) in order to estimate the quality parameters and tonnage of coal deposits.


Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References


28. Modis, K.; Sideri, D. Spatiotemporal estimation of land subsidence and ground water level decline in West Thessaly basin, Greece. *Nat. Hazards* 2015, 76, 939–954. [CrossRef]