Coal Segregation Control for Meeting Homogeneity Standards

Z. Huang, R. Kumar    J. Yingling, J. Sottile
Dept. of Elec. & Comp. Eng.    Dept. of Mining Eng.
University of Kentucky
Lexington, KY 40506
Email: zhdhuang, kumar, jyinglin, jsottile@engr.uky.edu

Abstract

In earlier works [1], [3], [4], [5] procedures were developed to control coal segregation to meet ash targets over large coal batches (e.g., a unit train of coal) while realizing high yields and economic savings. We have extended this work to address the issue of homogeneity control. The control objective is identical as before except that quality targets are required to be met over much smaller batch sizes. Two approaches are developed. The first approach assumes that quality variations are a manifestation of a locally stationary Gaussian process. Time-series-based approaches from the earlier work are used to forecast future quality levels as a function of time. Given this information, a chance-constrained mathematical programming formulation is used for making coal segregation decisions, where yield is maximized subject to a predefined limit on the risk of violating the quality target for the batch. An efficient solution procedure has been developed suitable for real time execution in segregation control. The second approach is simpler and is based on the principle of compensation. Here the “slack” amount of ash that may be added to the current batch without violating the target quality specification is continuously monitored. The measured ash values are compared to a threshold level based on the available slack and a simple statistical characterization of the quality variation process. Both approaches have been implemented in software and preliminary tests suggest that the second approach performs better than the first one, whereas both outperform the approach of earlier work.

1 Introduction

Coal segregation is an important mining procedure as it is a source of saving yield losses. On-line analyzers and computer control make it easier to segregate run-of-mine (r.o.m) coal. As shown in Figure 1, a particular block of r.o.m coal is physically segregated into two piles, one requiring no further washing as it meets the target quality specification, and
another requiring further cleaning as it doesn not meet the target quality specification. The segregation control is used to minimize the size of the wash pile, thereby maximizing the yield, while meeting the target quality specification for the coal of the no-wash pile. A typically used strategy is that, if the percentage ash of the newly arrived sample is above certain cutoff level, it is sent to the wash pile, and otherwise, it is sent to the no-wash pile. The cutoff levels are based on customer specifications and their practically implemented values are often determined by trial and error. Since the coal in the no-wash pile is sent to customer directly, it needs to meet quality specification.

2 Analysis for coal segregation

The decision to send a particular block of r.o.m coal on conveyor belt to wash or to no wash pile is a simple decision, but the issue of yield maximization (we want to send as much coal to no-wash pile as possible) makes it more involved.

The probability distribution of quality of r.o.m. coal is denoted by the density function \( f_Z(z) \), and its typical bell-shape nature is shown in Figure 2. The location of the mean of the distribution is shown as \( \mu_g \). The target specified by the customer is \( \mu^* \), and this target must be met on average by the coal in the no-wash pile. Normally, the customer specification is higher than the mean of r.o.m..

There are infinitely many ways to meet the quality specification, yet the yields are very different. Two possible methods, resulting in distributions shown as dashed lines are depicted...
Time-series and moving window approaches for forecasting the mean and variance of the distribution, assuming it to be locally stationary and Gaussian, were presented in [1],[3]. Forecasts from time series models, at various lead times ahead, were used to determine the cutoff ash level in time-series method; while in moving window method, a fixed number of samples, called the window width, in the immediate past were used to determine the cut-off ash value.

Two new approaches for meeting target quality specification for small batches have been developed, one is based on chance constrained, and another is based on principle of compensation. These two approaches are described below.

3 Chance constrained based approach

In order to formulate the chance constrained approach, consider the diagram of Figure 4. There are \( n \) samples of coal on the conveyor. The amount of coal in each sample equals the amount that passes by the ash quality analyzer between the times when two measurements are taken. The quality of \( n \) samples can be represented as \( Z = [z_1, \ldots, z_n]^T \in \mathcal{R}^n \). The first \( k \) samples, \( Z_1 = [z_1, \cdots, z_k]^T \), are detected samples
whose ash content values are already known, whereas the remaining $n - k$ samples, $Z_2 = [z_{k+1}, \ldots, z_n]^T$, are samples whose quality values are unknown.

The problem of coal segregation for homogeneous quality control (i.e., quality control over batches of small sizes) can be described as follows: each sample needs to be sent to either wash pile or no-wash pile, while the average ash content in the no-wash pile must meet target quality $\mu^*$, and the size of the no-wash pile is as large as possible, i.e., the yield, defined as the ratio of the no-wash pile to the total size of the two piles, is maximized. We let the decision variable $y_i = 1$ if and only if the $i$th sample is sent to the no-wash pile. Thus decision variable vector is given by $Y = [y_1, \ldots, y_n]^T \in \{0, 1\}^n$.

![Quality Analyzer](image)

**Figure 4:** Abstraction of coal segregation using on-line ash analyzer.

Next we describe the chance constrained approach for maximizing the yield while satisfying the customer specification in the no-wash pile for a pre-specified batch size of $r$ samples of no-wash pile, assuming that quality variations are a manifestation of a locally stationary Gaussian process.

The batch size of no-wash pile is required to contain $r$ samples. Let $m$ be the number of samples not washed thus far (note that at the initialization, $m = 0$); $\bar{\mu}$ be the mean of those $m$ samples; and $\mu^*$ be the average customer specification. The chance constrained formulation is next given:

$$\max_Y \left( \frac{\sum_i y_i}{n} \right)$$

subject to

$$\sum_i y_i = r - m \quad (1)$$

$$Pr \left( \frac{m\bar{\mu} + Y^T Z}{r} \leq \mu^* \right) \geq \alpha, \quad (\alpha \text{ represents a desired confidence level, say 99%}) \quad (2)$$

Here $m\bar{\mu} + Y^T Z$ is the total ash content of the no-wash pile under the choice of the decision variable $Y$. Under such a choice of $Y$, the no-wash pile will contain a total of $r$ samples, but at present the no-wash pile contains $m < r$ samples with an average ash content of $\bar{\mu}$. 
Note that in view of (1), the objective function can be rewritten as:

\[
\max \frac{\sum_i y_i}{n}
\]

\[\Leftrightarrow \min \frac{n}{\sum_i y_i}
\]

\[\Leftrightarrow \min \frac{\sum_i y_i + \sum_i (1 - y_i)}{\sum_i y_i}
\]

\[\Leftrightarrow \min Y \frac{(r - m) + \sum_i (1 - y_i)}{r - m}
\]

\[\Leftrightarrow \min Y \sum_i (1 - y_i)
\]

In other words, we must decide for each sample \( i \in \{1, \ldots, n\} \) whether or not to wash it so that:

1. The total number of samples sent to the no-wash pile based on this decision is \( r - m \) (batch size minus the current size of the no-wash pile);

2. Probability that the specification is met exceeds a desired confidence level \( \alpha \) (typically 99%);

3. The yield, defined as the ratio of samples sent to the no-wash pile to the total number of samples examined, is maximized.

In the following we show how the above chance constrained problem can be reduced to an instance of deterministic integer program under the assumption that the unobserved part of \( Z \) is a Gaussian random variable. We first separate \( Y^T Z \) into the deterministic and the random part, i.e., \( Y^T Z = Y_1^T Z_1 + Y_2^T Z_2 \), where \( Z_1 \) and \( Z_2 \) are as defined above, and \( Y_1 \) is the sub-vector of \( Y \) consisting of its initial \( k \) entries, whereas \( Y_2 \) is the sub-vector of \( Y \) consisting of the remaining \( n - k \) entries.

Under the assumption of Gaussianness, we have \( Z_2 \sim N(\mu, \Sigma) \) which implies:

\[
Y_2^T Z_2 \sim N(Y_2^T \mu, Y_2^T \Sigma Y_2)
\]

\[\Rightarrow \frac{Y_2^T (Z_2 - \mu)}{\sqrt{Y_2^T \Sigma Y_2}} \sim N(0, 1)
\]

Next we have the following equivalence,

\[
[Y_2^T Z_2 \leq r \mu^* - m \bar{\mu} - Y_1^T Z_1]
\]
\[
\left[ \frac{Y_2^T(Z_2 - \mu)}{\sqrt{Y_2^T \Sigma Y_2}} \leq \frac{r \mu^* - m \bar{\mu} - Y_1^T Z_1 - Y_2^T \mu}{\sqrt{Y_2^T \Sigma Y_2}} \right]
\]

From the above equivalence it follows that

\[
Pr \left[ \frac{m \bar{\mu} + Y^T Z}{r} \leq \mu^* \right] \geq \alpha
\]

\[
\Leftrightarrow Pr \left[ Y_1^T Z_1 + Y_2^T Z_2 \leq r \mu^* - m \bar{\mu} \right] \geq \alpha
\]

\[
\Leftrightarrow Pr \left[ Y_2^T Z_2 \leq r \mu^* - m \bar{\mu} - Y_1^T Z_1 \right] \geq \alpha
\]

\[
\Leftrightarrow Pr \left[ \frac{Y_2^T(Z_2 - \mu)}{\sqrt{Y_2^T \Sigma Y_2}} \leq \frac{r \mu^* - m \bar{\mu} - Y_1^T Z_1 - Y_2^T \mu}{\sqrt{Y_2^T \Sigma Y_2}} \right] \geq \alpha
\] (3)

For \(X \sim N(0,1)\), let \(Q(x) = Pr[X \geq x]\), or equivalently, \(Pr[X \leq x] = 1 - Q(x)\). Then, since \(\frac{Y_2^T(Z_2 - \mu)}{\sqrt{Y_2^T \Sigma Y_2}} \sim N(0,1)\), Equation (3) can be rewritten as:

\[
1 - Q\left( \frac{r \mu^* - m \bar{\mu} - Y_1^T Z_1 - Y_2^T \mu}{\sqrt{Y_2^T \Sigma Y_2}} \right) \geq \alpha
\]

\[
\Leftrightarrow Q\left( \frac{r \mu^* - m \bar{\mu} - Y_1^T Z_1 - Y_2^T \mu}{\sqrt{Y_2^T \Sigma Y_2}} \right) \leq 1 - \alpha
\]

Since \(Q\) is a decreasing function of its argument, this becomes

\[
\Leftrightarrow \frac{r \mu^* - m \bar{\mu} - Y_1^T Z_1 - Y_2^T \mu}{\sqrt{Y_2^T \Sigma Y_2}} \geq Q^{-1}(1 - \alpha)
\]

\[
\Leftrightarrow Q^{-1}(1 - \alpha)\sqrt{Y_2^T \Sigma Y_2} + Y_1^T Z_1 + Y_2^T \mu \leq r \mu^* - m \bar{\mu}
\] (4)

Thus we are able to replace the chance constraint (2) by an equivalent deterministic constraint (4). We are dealing with a linear objective function subject to two constraints, one of which is linear (constraint (1)), and the other quadratic (constraint (4)). The optimal solution is computationally hard for such problems. So we next present an efficient search technique for computing a near-optimal solution.

- First note that the yield is maximum when none of the samples is sent to the wash pile, i.e., \(y_i = 1\), for all \(i\).

- So we set \(y_i = 1\) for all \(i\) (which implies \(n = r - m\)) and verify (4), noting that when \(y_i = 1\) for all \(i\), \(Y_2^T \Sigma Y_2, Y_1^T Z_1, Y_2^T \mu\) equals the sum of entries in \(\Sigma, Z_1, \mu\) respectively.
If (4) holds, then an optimal is found, else we increment \( n \) by 1, setting it to \( r - m + 1 \), in which case, \( y_i = 0 \) for exactly one \( i \).

- Since \( Y \) must be chosen so that LHS of (4) is smallest possible, \( y_i \) is set to zero for that particular \( i \) for which the sum over the \( ith \) entry of \( Z_1 \), the \( ith \) entry of \( \mu \), and the sum of \( ith \) column and \( ith \) row of the covariance matrix \( \Sigma \) is maximum.

- Verify (4) under this selection of \( Y \), and if it holds, then an optimal is found; else we increment \( n \) by 1, setting it to \( r - m + 2 \), i.e., \( y_i = 0 \) for two \( i \)'s, etc.

- We continue in the above manner till an optimal is found or the limit for local stationarity conditions is reached, in which case no optimal solution can be found.

For the unobserved variable \( Z_2 \), we use ARIMA time-series model to capture the stochastic characteristics of the fluctuations in the coal-quality levels over time and get its forecast values using this model. They are two major kinds of time series: stationary and nonstationary. In the former case, both the mean and variance of the values remain stable over time. In the later case, the mean or the variance, or both, change with time. If nonstationary, stationarity can often be achieved by dealing with a difference process. A low order differencing generally produces a stationary time series. This happens to be the case of coal quality observed at the mine.

So a basic assumption is that the observed time series of coal quality can be modeled as the output of a linear filter subjected to a input white-noise process consisting of a number of autoregressive or moving average terms [2]. These models explicitly characterize the correlations among consecutive samples. Moreover, the model may be used to forecast future values of \( z(t) \), given the process history.

Several data sets containing r.o.m. ash values were obtained from an operational mine. The quality analyzer in this mine samples the r.o.m. coal moving on a conveyor belt. Percentage ash is one of the quality values determined by the analyzer. These ash values, obtained at 5-sec intervals, were modeled as time series. These models were then used to forecast, obtaining both the mean and variance of \( Z_2 = [z_{k+1}, \ldots, z_n]^T \).

The whole procedure includes 3 steps. Firstly, we need to judge whether our series is stationary. If it is not, then we need to do differential transformation or power transformation to get a stationary series. Further information is given in [6]. After the stationary series is obtained, we obtain a suitable model fitting the correlation values to get the needed parameters of the ARIMA model. Finally, we check whether the model is valid. Residual test is most commonly used for this purpose [2]. If the validity is not verified, then we change the selected model and repeat the steps until a valid model is found.

A comparison of the performance for different batch sizes is shown in Figure 5. We can see from this figure that it has better yield than the naive method, which implements customer specification directly as the cutoff value. As the batch size increases, the yield also increases. Our algorithm can be implemented for large batch sizes since a large batch size can be regarded as a collection of many small batch sizes.
Figure 5: Performance of Chance Constrained Approach.
4 Compensation Based Approach

A plausible optimal method of segregation is shown in Figure 6, where all samples with ash values below a particular cutoff value, denoted $Z_c$ in the figure, are accepted, and all those above the cutoff are rejected.

![Figure 6: The optimal method of segregation.](image)

The problem with such an approach is the optimal cutoff value is unknown since the distribution of $Z$, $f_Z(z)$, is unknown a priori. So we propose a compensation based approach. Note that there is a certain “slack” present in the no-wash pile given as the sum, taken over all samples, of the difference between the target specification and sample’s ash value. We can make use of such slack to accept samples whose ash content is higher than the target specification by an amount equal to the current slack, whenever it is possible. However, in order to make an efficient use of the slack, we don’t accept the samples whose ash content is too high, which is determined to be a sensibly chosen upper bound, $(m + \sigma)$, where the parameters $m$ and $\sigma$ are chosen using one of the four methods described below.

Figure 7 illustrates the algorithm for the compensation based approach. The ash content of a newly arrived sample, denoted $Z_{new}$, is compared against the target specification ($\mu^*$) plus the present slack, and if $Z_{new} \leq \mu^* + \text{slack}$, and also $Z_{new} \leq m + \sigma$, then the newly arrived sample is sent to the no-wash pile. Here the parameters $m$ and $\sigma$ are used to upper bound the ash content of the sample sent to the no-wash pile. The slack value is initially set to zero, and it is revised as follows:

$$\text{slack} = \begin{cases} \text{slack} + (\mu^* - Z_{new}) & \text{if newly arrived sample sent to no-wash pile} \\ \text{slack} & \text{otherwise} \end{cases}$$

Note that if $m + \sigma$ is very large, then a newly arrived sample is sent to the no-wash pile if and only if $Z_{new} \leq \mu^* + \text{slack}$. Thus if $Z_{new} = \mu^* + \text{slack}$, then the revised slack will become zero, causing the next sample to be sent to the no-wash pile if and only if it meets the target specification. This may not be optimal for yield maximization, and so, we choose an appropriate value for $m + \sigma$, using one of the four methods given below.

- **Method 1**: $m = \text{mean of all past samples, } \sigma = \text{SD of all past samples.}$
Figure 7: Principle of Compensation Based Approach.
Method 2: \( m = \text{mean of forecast samples}, \sigma = SD \) of forecast samples.

Method 3: \( m = \text{mean of most recent } N \text{ samples (we use } N = 50), \sigma = SD \) of most recent \( N \) samples.

Method 4: \( m = \mu^*, \sigma = \text{average deviation from } \mu^* \) of past samples in the wash pile.

All four methods work well as shown in Figure 8; their yield depends on the series stochastic characteristics. Usually, Method 4 provides a better yield.

![Figure 8: Comparison of Various Compensation Principle Based Methods.](image)

We can slightly change our compensation approach and get a modified compensation method, shown in Figure 9, where two adjustable parameters \( R_1 \) and \( R_2 \) are added to adjust the threshold, which is now computed as \((1 - R_1)m_1 + (1 - R_2)\sigma\). In Figure 7, \( R_1 \) and \( R_2 \) are simply zero.
Figure 9: Modified Compensation Based Approach.
In Figure 10, we choose $m = \mu^*$, $\sigma$ = average deviation from $\mu^*$ of past samples in the wash pile as in Method 4 above, since it works the best among the four methods, and choose different values of $R_1$ and $R_2$.

Suitable $R_1$ and $R_2$ can increase the yield, but high $R_1$ or $R_2$ may violate the target specification although increasing the yield. At the beginning of segregation, $R_1$ can be 0.1 or 0.2. High positive slack means that $R_2$ can be increased to raise the yield. Negative slack means that $R_1$ should be decreased.

![Figure 10: Comparison of the performances of modified compensation method.](image)

5 Conclusion and Discussion

Physical segregation of coal may result in dramatically better economies for the mine operators. The potential benefits of on-line ash analyzer is exploited by proposing high yield
and efficient yet simple to implement on-line segregation control techniques. Two possible segregation techniques, chance constrained and compensation, by which optimal control of segregation can be achieved, are presented. Both methods can be applied for larger batch sizes.

This paper demonstrates the viability of the two proposed techniques. The chance constrained approach searched among detected samples and forecast samples. It can be improved by extending the search scope. The compensation approach achieved a higher yield. The compensation approach has also the advantage of being simple and a faster computation time, making it suitable for on-line implementation. Compensation approach can meet quality target while achieving higher yield by suitable selections of $m$ and $\sigma$. Modified compensation method makes the algorithm flexible to operators by allowing the choice for additional parameters $R_1$ and $R_2$.

An important advantage of these techniques is that they are algorithmic and hence require no special purpose hardware for implementation. All that needs to be done to implement them at a mine is to modify the existing software that controls the flop gates. The authors have written such code in C and Gauss that takes quality values from the ash analyzer as input, and outputs the flop gate control decisions.

**Acknowledgement**

The authors acknowledge DOE, KY DOE-EPSCOR for providing partial financial support for the research and the person at the cooperating mine for their excellent assistance in obtaining the data necessary for this study.

**References**


