Monitor-While-Drilling - based estimation of rock mass rating with computational intelligence: the case of tunnel excavation front☆

Galende-Hernández, M.[a], Menéndez, M.[c], Fuente, M.J.[b], Sainz-Palmero, G.I.[b,*]

[a] *CARTIF Centro Tecnológico, Parque Tecnolgico de Boecillo, 47151 Boecillo (Valladolid), Spain.*
[b] *Dpt. of Systems Engineering and Control. School of Industrial Engineering. University of Valladolid. 47011 Valladolid, Spain.*
[c] *VIAS y Construcciones, C/ Orense 11, 28020 Madrid, Spain*

Abstract

The construction of tunnels has serious geomechanical uncertainties involving matters of both safety and budget. Nowadays, modern machinery gathers very useful information about the drilling process: the so-called Monitor While Drilling (MWD) data. So, one challenge is to provide support for the tunnel construction based on this *on-site* data .

Here, an MWD based methodology to support tunnel construction is introduced: a Rock Mass Rating (RMR) estimation is provided by an MWD rocky based characterization of the excavation front and expert knowledge.

Well-known machine learning (ML) and computational intelligence (CI) techniques are used. In addition, a collectible and "*interpretable*" base of knowledge is obtained, linking MWD characterized excavation fronts and RMR.

The results from a real tunnel case show a good and serviceable performance: the accuracy of the RMR estimations is high, $Error_{test} \cong 3\%$, using a generated knowledge base of 15 fuzzy rules, 3 linguistic variables and 3 linguistic terms.

☆This work has been partially supported by the Spanish Ministry of Economy and Competitiveness and the European Regional Development Fund (FEDER) through the Project no. DPI2015-67341-C2-2-R

*Corresponding author: Sainz-Palmero, G. I., (gresai@eii.uva.es)

This proposal is, however, is open to new algorithms to reinforce its performance.

Keywords:
Tunnels, MWD, FRBS, Selection/Extraction of Features, Clustering, RMR, Decision Making

---

1. Introduction

Nowadays, most operations and tasks in modern industrial activities are monitorized and, in general, the data and performance are logged for different reasons, such as maintenance, safety, etc. Construction, mining, and the tunnelling industry involve activities, equipment and technologies that can match the current trend concerning the capture and logging of data to take advantage of the embedded information for a better performance and improvement of the processes. So, the challenge is how to evaluate all the available data, information and expert knowledge to improve these processes from all points of view: safety, management, quality, etc. which can all lead, of course, to a more profitable business.

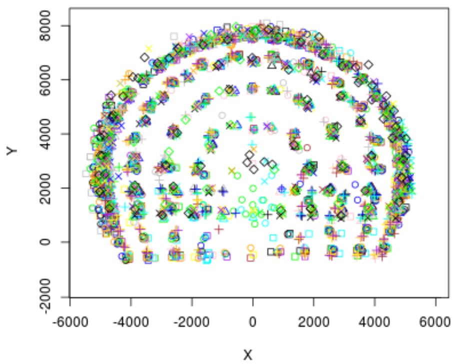
The proposal of this work is focused on the tunnelling industry, to be precise the specific case of railway tunnels, but this can be applied to other similar cases such as road tunnels; underground mining and utilities, etc. Tunnel excavation has used two main methods: *Drill & Blast* and *Tunnel Boring Machine (TBM)* [1]. The first is the most popular excavation method for conventional tunnelling, in particular for railtrack tunnels. In any case, both methodologies involve the use of computer and control based machinery to capture and log data of different natures concerning the process: this is the so-called *Monitor or Measurement While Drilling* (MWD) data [2].

Conventional tunnels can be defined as the construction of underground openings of any shape by a cyclic construction process [1]. This type of tunnels are usually made using the *Drill & Blast* excavation method. Here, the "*jumbos*" are the machinery for face drilling (See Figure 1): drilling several holes in the rock wall face area, known as the excavation front, after which these holes are filled with explosives. The explosion causes the collapse of the rock and the lengthening of the tunnel. Then, a new excavation front appears ahead, and this cycle is repeated.

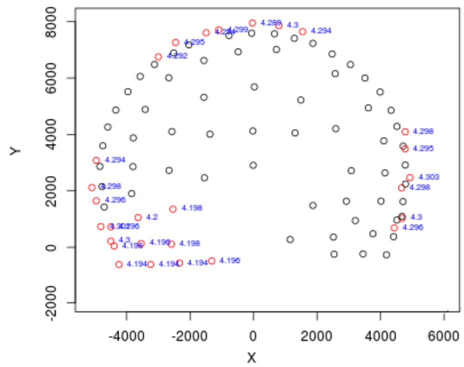
The result of each of these steps is an excavation front, and its performance is based on such measurements as the rock mass stability, or other

(a) Jumbo and excavation front

(b) Jumbo drilling holes on the excavation front



(c) Overlapped patterns of drillings of the excavation fronts



(d) Pattern of drilling holes for an excavation front

Figure 1: Drill & Blast excavation method.

tunnel design indexes that may be available. Here, the RMR system is a well-known geomechanical classification system for rocks [3] which is very popular for tunnel construction and is used, in its several versions, as a design tool to determine a tunnel support type. It is a vital element for this kind of business.

All of this is carried out by the technicians in charge of these operations. In this context, some extra support for the conditioning of the excavation front would be very interesting in order to deal with the uncertainties associated with the problem and its implications: safety, management, budget, etc. So far, one of the most usual MWD based applications has been the guiding, positioning and conditioning of the machines, but other valuable and serviceable support could be possible. How this data can be used to give support on-site to improve the efficiency of the tunnelling process is still an open and debatable issue.

Here, the proposed methodology gives *"on-site"* support to the usual work flow of this below-surface construction based on the MWD data recorded in real time and ML & CI techniques: *a priori* RMR estimation of the next excavation front to be discovered is provided using an unsupervised MWD based characterization of the drilling front and the available expert knowledge. This characterization is supported by a clustering of MWD based drillings into a few rocky categories. In order to deal with this latter task, an unsupervised selection of the most relevant MWD variables is made. Both previous steps also address the almost unavoidable issue of the reduction in complexity, or dimensionality, regarding the information managed so that it can be used by ML& CI techniques and the users. This dimensionality challenge is critical: every excavation front is characterized, and summarized, to estimate the RMR using its own hole drillings, which, due to the high number of them for each front, are summarized by a few MWD drilling based rocky clusters. On the other hand, all this is supported by MWD data containing a high number of variables, so an analysis and selection of these data variables is compulsory.

All these previous proposal steps are implemented by several different ML & CI algorithms for each one, in order to achieve a better performance. This means carrying out user-tuned decision making at each stage to get the stage output. Which is also made by fuzzy approaches for a better robustness and , as well as using linguistic terms during this decision making. Finally, the RMR estimation is made using linguistic and scatter genetic fuzzy rule based systems (FRBS). During this final task, the expert knowledge is supplemented by the unsupervised MWD knowledge from the previous steps. Additionally, this last stage allows a collectible knowledge base to be obtained through by fuzzy rules linking MWD rocky based excavation fronts with RMR estimations.

This entire methodology implies that data, expert knowledge and information must be available in sufficient quantity and quality to be implemented using these ML & CI approaches, so as to be able to reach a sufficiently good performance for the desired goals.

The rest of this paper is organized as follows: in Section 2, a brief review of related works is done. Section 3 summarizes the computational intelligence for this proposal. The proposed methodology is detailed in Section 4. Section 5 explains and discusses the experimental work and its results. Finally, the conclusions of this work are set out in Section 6.

2. Related Works

Tunnelling, and all works involving geological mechanics, have a high level of uncertainties concerning the geomechanics and their evaluation. Some popular views to address these uncertainties are based on soft computing, artificial intelligence, computational intelligence, or machine learning, etc. All these well-known approaches can deal with this challenge: fuzzy logic permits fuzzy information to be dealt with; artificial neural networks allow us to learn from collected data; a genetic algorithm is able to optimize parameters, while fuzzy rule based systems describe the knowledge known or learnt by "if-then" rules, providing data-driven or knowledge based models that permit estimations of the geomechanical evaluations. A general review of this can be found in [4], where the authors set out a general review of soft computing approaches for mining problems.

Neural network based approaches can be found in [5], where a three-layered Feed-Forward Backpropagation network is used to predict the stress-strain response of intact and jointed rocks using data reported in the literature. In [6] and [7], a preliminary application of backpropagation neural networks is considered to optimize the mine support parameters using simulation data and for analyzing rock mass parameters in tunnelling. MultiLayer Perceptron (MLP) and Radial Base Functions (RBF) artificial neural networks are used in [8] to predict rock properties from sound levels generated during in-lab drilling. Neural Networks and data mining are used in [9] for tunnel support stability using a very high number of inputs based on off-line geological measurements, but here the well-known curse of dimensionality can influence its performance. The Support Vector Machine (SVM) and neural networks are used to predict tunnel convergence in [10], but any true variable selection is carried out for the best inputs. In [11], however, artificial neural networks and Hidden Markov Models (HMM) are used to provide a probability model of the geology class.

Fuzzy logic principles are used by [12] to generate a Mamdani fuzzy rule based system, using Mamdani fuzzy "if-then" rules, to be applied to the Geological Strength Index (GSI) and tuned by the intuition method, though its application in comparison with data driven models would seem to be debatable. In most cases, Fuzzy Logic is hybridized with artificial neural networks, resulting in the neuro-fuzzy approach, which is one of the most popular approaches for dealing with uncertainty when predicting and estimating in geotechnical engineering: [13] reviewed of fuzzy applications, where ANFIS is the most popular neuro-fuzzy implementation when there are other, more powerful, neuro-fuzzy approaches with extra, and even better, performance.

In [14], a neuro-fuzzy Takagi-Sugeno system is used for rock cutting in mining machines with simulated data; while in [15], models of rock fragmentation are obtained by SVM and ANFIS neuro-fuzzy systems, with a debatable input selection by Principal Component Analysis (PCA). A prediction of the advance rate of tunneling is carried out in [16] based on ANFIS and using off-line heterogeneous information as input.

On the other hand, the RMR system [3] is very popular in tunnel constructions. This RMR rating has been predicted in previous, similar approaches: in [17], where an RMR rating is predicted in simulation with its own RMR parameters (weightings) as inputs, using an ANFIS system and a Fuzzy c-means algorithm. In [18], however, a multivariable regression provides a predictive equation to be compared with an FRBS with a very large base of fuzzy rules, whose tuning is not specified, using the RMR weightings as inputs. In addition, [19] carries out an RMR prediction using a chaos-ANFIS model and continuous functions for RMR weightings, while [20] involves an ANFIS model and a genetic algorithm for estimating the deformation modulus and the RMR system related to other geomechanical indexes.

Data mining techniques are applied to the geomechanical characterization of rock masses, predicting the deformation modulus in [21] based on the depth, the weightings of the RMR systems, etc. Moreover, [22] uses multivariate linear, non-linear and polynomial regression analysis of RMR input parameters and TBM field penetration indexes to improve the hard rock TBM; while in [23], the rock deformation modulus is based on polynomial and multiple regression analysis of the RMR systems.

A critical issue for all these approaches is the availability of the data and expert knowledge. In current tunnel excavations and mining, the MWD data is an essential information source about the work in progress, while the modern tunnelling facilities provide MWD data. In [2], an overview of MWD techniques and their scope in the excavation industry is presented. Then, one of the open challenges is how this MWD data could be used to provide support. Concerning this challenge, [24] identifies potential relations between parameters captured in conventional MWD applications with critical rock and rock mass properties in blast and underground designs.

Pattern recognition and fuzzy techniques are considered in [25], where a rock mass classification is managed as a multi-feature pattern recognition problem based on RMR parameters, or on MWD data in [26]. Also, in [27], an academic study can be found involving MWD data and backpropagation neural networks for supervised learning. A discussion about machine learning approaches for geosciences and remote sensing can be found in [28]. [29] develops research into pattern recognition approaches for MWD based rock identification using SOM and backpropagation approaches with different inputs.

Another view can be found in [30], where the authors address the problem of MWD feature selection for automatic rock recognition. The selection is based on MWD drilling data measurements on the frequency domain. The same authors, in [31], focus on automatic rock recognition from MWD drilling performance data by unsupervised clustering. Finally, [32] uses MWD for the detection of coal seams without geophysical data.

Summarizing, the ML & CI approaches have been used in this domain in many ways, most of them to address specific challenges in geological issues: regarding the use of geological features to obtain estimations of other rock characteristics, recognizing rock types, estimating slope stability, maximum charges or soil failures, links between geological features, etc. In most cases, raw machinery/sensor data are not used as inputs, whereas geological characteristics or weightings of the RMR systems are used as the inputs. On the other hand, very few cases involve RMR and/or MWD data. Where RMR is estimated using partial RMR parameters and MWD data to link geological conditions, in general, many partial challenges are faced. However, in may cases, the experimental methodology is not well detailed, so it makes checking the proposals difficult.

Nevertheless, our proposal provides an open general, and methodological, approach covering from raw machinery data on a sub-symbolic level, an on-site estimation of a critical service parameter (RMR value) for the next excavation front on a symbolic level of knowledge. This is based on MWD data from the drilling rig and ML & CI techniques , which additionally provides a rocky characterization of the excavation front, as well as a collectible and ''*interpretable*'' knowledge base linking the MWD rocky characterization of fronts with RMR values. This high level of on-site support means relevant resource savings and better safety.

3. Basis for a computational intelligence based tunnel excavation

In order to carry out this proposal concerning tunnel excavation, some well-known techniques from the fields of machine learning and computational intelligence are used to implement it. This involves techniques for variable selection, clustering based data modelling, fuzzy based modelling and decision making. These well-known techniques are very briefly described below.

The selection of variables is based on two reputed, unsupervised feature extraction techniques: Principal Components Analysis (PCA) and Factor Analysis (FA).

PCA (*Principal Components Analysis*) is a well-known statistical procedure, generally used as an unsupervised feature extraction technique for dimensionality reduction. In this way, PCA provides a set of new features, called Principal Components (PCs), as a linear combination of original variables [33], [34]. Non linear PCAs are also available [35], [36]. On the other hand, a PCA technique can be used as an unsupervised variable selection approach, such as in [34] and [37].

FA (*Factor Analysis*) is a regression technique which has some similarities with PCA. It expresses a set of available variables as a linear combination of a smaller number of other unobservable variables, called Factors, and some error terms [38].

In addition, some standard clustering techniques with different nature and performance are considered in order to obtain different views of the data managed. The clustering algorithms taken into account in this work are:

K-Means is a center-based clustering algorithm, which allows similar instances to be grouped into k clusters, minimizing the distance between cluster data and its center [39]. Due to the unsupervised nature of this process, the "optimal number" of clusters, k, is unknown. Here, the challenge is to find this number of clusters, which can be addressed using some validity criteria about the clustering quality, such as: Calinski-Harabasz CH [40]), Davies-Bouldin DB ([41], Dunn's D [42], SD Validity SD [43], Silhouette S [44] or Xie & Beni's XB [45]. The best cluster, or data partition, is one that maximizes CH, D and S indexes and minimizes DB, SD and XB.

X-Means is an extension of K-Means using Bayesian Information Criterion that is able to determine the optimal number of clusters [46].

DBSCAN is a density-based clustering algorithm, which checks the ϵ neighbourhood around each point and ensures at least $MinPts$ points per cluster [47].

Spectral is a clustering algorithm based on graphing Laplacians [48]. In order to set a scaling parameter σ to calculate the affinity matrix and optimal number k of clusters for K-Means, previous validity criteria such as CH, DB, D, SD, S and XB can be used.

FCM (*Fuzzy C-Means*) is a fuzzy clustering algorithm similar to K-Means, involving "fuzzy" concepts. The validation criteria to determine the number of clusters are: Xie & Beni's XB [45], Fukuyama-Sugeno FS [49], Partition Coefficient PC [50] or Partition Entropy PE [51]. In this case, PC must be maximized while XB, FS, PE must be minimized.

The final target is to make predictions based on the previous techniques and data. This is based on two genetic fuzzy systems, S-IRL and L-IRL, with different fuzzy performance, permitting the unsupervised knowledge and expert knowledge to be gathered together, due to its supervised nature. On the other hand, the fuzzy nature of both algorithms implies a capability of explanation by fuzzy rules of the learnt knowledge base, even in linguistic terms, that gives an extra performance:

S-IRL (*Scatter Iterative Rule Learning*) is a modeling algorithm that, guided by a genetic algorithm, is able to generate scatter fuzzy rule based systems [52]. The number of linguistic terms (nLT), along with other options, must be defined *a priori*.

L-IRL (*Linguistic Iterative Rule Learning*) is a modeling algorithm that, following a similar strategy to S-IRL, is able to generate linguistic fuzzy rule based systems ([53]).

In this proposal, variable selection, clustering or the prediction tasks are based on different algorithms, with different performances and results, even controversial, between algorithms for the same targets. So, a decision-making is essential to obtain a result considering all these available alternative results. Here, the decision making is based on the aggregation function, the RIM quantifier and the OWA operators:

OWA (*Ordered Weighted Average*) is an aggregation operator used to obtain a single representative value from others according to a weighting vector [54]. These weights are usually computed through a Regular Increasing monotone (RIM) Quantifier, which introduces *andness* and *orness* measures that can implement *linguistic terms*.

4. Computational Intelligence for MWD and RMR: an approach

The current proposal is based on the application of well-known computational intelligence techniques to an engineering problem concerning how available MWD data can be used during tunnel excavation to give high level support. The issues of this approach are based on ML & CI techniques, involving from data analysis to prediction and decision making, as well as the extraction of a knowledge base. Clearly and methodologically defining the stages to be dealt with in order to take advantage of the MWD data using ML & CI for the estimation of operational tunnel parameters, and how every stage can be implemented by these techniques.

The proposal has a double point of view:

1. First, this proposal introduces a general, and open, methodology to take advantage of available MWD data and collected expert knowledge, involving standard ML & CI algorithms that cover every stage from data processing to the prediction of design parameters. These predictions take advantage of the MWD data in real time, evaluating in an unsupervised way the best variables of the prediction, as well as providing a way to summarize the dimensional complexity of the challenge. On the other hand, the methodology is open to incorporating or changing new CI approaches that can improve the performance of the system from different points of view, and also tuning parameters.
2. Second, based on this methodology, an approach for making predictions of the RMR values for excavation fronts are set out, but other tunnel designed parameters can also be considered, applying the same methodology:
    - An unsupervised MWD based selection of variables is carried out to find the most valuable.
    - The drillings described by the MWD selected variables are clustered into categories.
    - Based on these drilling categories, every excavation front is characterized and summarized by the distribution of a few MWD Rocky based drilling categories.
    - Genetic FRBS approaches forecast the RMR values for every excavation front, based on its MWD Rocky based drilling characterization and the collected expert knowledge concerning RMR.

Here, the expert knowledge is incorporated due to the supervised nature of the genetic fuzzy systems considered in this work, and a new and collectible knowledge base is learnt and available by fuzzy rules.

- Each of these steps is faced by different ML & CI approaches, every one contributing with its own criterion for the goal, which imply alternative solutions and performances. In order to take into account all these alternative criteria, a decision making must be carried out based on these alternative criteria, setting out a rank of approaches for each goal.

In Figure 2, the general scheme of this methodology, as well as its key issues, are described. The major goal of this work is the prediction of tunnel design parameters, the current version concerns the RMR:

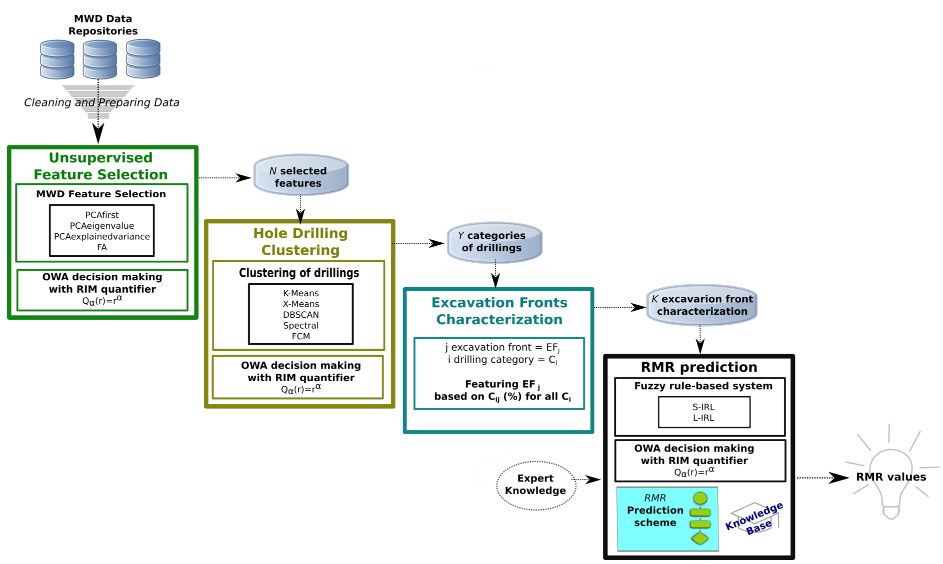


Figure 2: Main stages of the Proposal

1. **Stage 1 -** *Unsupervised Variable Selection:*, the MWD data available is processed in order to validate data, removing outliers, fixing missing values, etc. Then a selection of the most relevant variables is carried out based on two well-known unsupervised techniques: PCA and FA. Different numbers of factors and principal components are considered as criteria for variable evaluation. Finally, an OWA based multicriteria

decision making is carried out to obtain an unsupervised ranking with the most relevant MWD variables to be used in the following stages.

2. **Stage 2 -** *Hole Drilling Clustering:*, taking into account the previous MWD variable selection, different clustering algorithms and indexes are used to validate each alternative clustering or partitioning. Different categories of hole drillings are obtained and described based on the MWD variables. Once again, based on alternative clustering categories and performances, an OWA based multicriteria decision making is carried out to provide a ranking of the parameters for the clustering algorithms involved.
3. **Stage 3 -***Characterization of Excavation Fronts:*, based on the hole drilling categories obtained in the previous stage, every excavation front is characterized and summarized through its own distribution of MWD based hole drilling categories.
4. **Stage 4 -** *Prediction of RMR Values and Weightings:*, based on linguistic and scatter fuzzy systems, as well as the expert knowledge collected, the prediction of the RMR values are set out. The best prediction model is based on an OWA decision making applied over all the alternative fuzzy models and their performances from different points of view, such as error, complexity or number of linguistic variables and terms. Besides, the fuzzy nature of these algorithms permits a knowledge base to be obtained, which is expressed by (linguistic) fuzzy rules, linking the MWD data with the expert knowledge available.

Stages 1 and 2 are essential to enabling the rest of the steps in the methodology to be addressed: determining what MWD variables are relevant in order to permit a hole drilling based clustering without the problems associated with the high dimensionality of the input space. Then, this MWD drilling based clustering permits Stage 3, summarizing the excavation front into a few rocky clusters, which once more avoids the dimensionality issues in Stage 4, thus permitting the RMR prediction based on FRBSs in the right conditions to be dealth with.

The *on-site* predictions and in advance, of the RMR values are extremely valuable for the technicians to manage uncertainties and plan the pattern of hole drillings for the new excavation front ahead, as well as the support needed for the tunnel walls. These estimations allow extra support to minimize risks in the advance of the tunnel, which not only concerns safety issues, but also managing, planning and economic issues. On the other hand, intermediate results and issues of this proposal permit us to know the most relevant variables from the MWD data set, in order to characterize the different drilling types and excavation front, which are described by the MWD based variable and can be linked with geological knowledge. The genetic fuzzy approaches used to predict RMR values permit a knowledge base, expressed by fuzzy rules and linguistic terms, concerning the MWD variables and expert knowledge regarding the RMR to be obtained.

5. Experimental Work

In order to implement and check the previous proposal, a case study based on a real railtrack tunnel excavation was used: this excavation concerns 2 tunnels, one for each direction. The methodology of excavation used was *Drill & Blast*.

When each excavation front was available, some design parameters concerning the tunnel were estimated by technicians, such as RMR values to evaluate the mass rock stability, usually estimated by geologists according to some protocols and their own experience. Based on these values, some tasks were decided, such as the patterns for the next hole drillings or the primary support for the tunnel. The number of available excavation fronts were 52 for both tunnels, and 3640 hole drillings made for them. For every excavation front (see Table 1), around 75 hole drillings were made, which had an average depth of around 3 meters, and the drillings took around 50 seconds.

Table 1: Available Data Sets

Number of:	
Excavation Fronts	52
Hole Drillings	3640
Samples	16701
MWD Variables	10

Table 2 shows the 10 MWD variables captured by the jumbo machine during the excavation drillings. After a data cleaning process, 3551 drillings were available: missing values, out of range, or duplicate drillings were removed from the database. In order to work with the variable records of each hole drilling, many alternatives were attempted. Finally, each drilling is managed as a function of the drilling advance on time and depth, then the

Fourier Transform (FT) is applied for both cases. This FT transformation is featured for each MWD variable by the Amplitude to $\omega = 0$ (A_0), its band width (B_W) and the difference of amplitudes between the two first fundamental frequencies (R_A). So, for each variable from Table 2 concerning a drilling, the FT considering time (T) and space (HD) are computed through three values: A_0, B_W, R_A. In short, each drilling is described by 16 Fourier transformed MWD variables, each one being featured and summarized by 3 FT values: A_0, B_W and R_A.

The following sections show in detail the methodologies used for each stage of this experimental work, and the analysis of these results.

5.1. Experimental Methodology

Algorithm 1 Variable Selection from MWD data

Require: Z-score standarization of characteristic values from Fourier Transform.
  for FeatureSelectionAlgorithm=PCA:FA **do**
    if PCA **then**
      l_{ij} =Perform PCA over covariances matrix
      for each variable v **do**
        $weight_1(v)$: Only first PC is considered
        $weight_2(v)$: PCs whose $eigenvalue$ is over $mean(eigenvalue)$ are considered
        $weight_3(v)$: PCs whose $\sum(ExplainedVariance(P_j)) \geq 90\%$ are considered
      end for
    else if FA **then**
      c_{ij} =Apply FA(number of factors according to Horn's Parallel Analysis, "maximum likelihood" estimation and "varimax" rotation)
      for each variable v **do**
        $weight_4(v)$: All factors are considered
      end for
    end if
    $weight$ are standardized and ranked with OWA
  end for
  Select variable with highest OWA rank

The first step is devoted to carrying out an unsupervised variable selection from the MWD processed and transformed data. The goal of this stage is to make a ranking of the most valuable variables in order to select the best ones, considering an unsupervised point of view. Four unsupervised alternatives have been applied to MWD variables: PCA considering 3 different criteria and FA. Each alternative supplies its own ranking: in order to set out a global ranking, an OWA based decision making is carried out. OWA weights are based on a RIM function such as $Q_\alpha(r) = r^\alpha$, with $orness(Q_\alpha) = \frac{1}{1+\alpha} = \{0.4, 0.5, 0.6\}$. This methodology is summarized in Algorithm 1.

Algorithm 2 MWD Data based Clustering for Characterization of the Excavation Fronts

Require: Selected features in Algorithm 1 and $[0, 1]$ range standardisation
  for ClusterAlgorithm=K-Means:X-Means:DBSCAN:Spectral:FCM **do**
    if $K - Means$ **or** FCM **then**
      for $NumberOfClusters$=2 **to** 12 **do**
        Run K-Means($NumberOfClusters$) or FCM($NumberOfClusters$)
        Calculate validity criteria
      end for
      Set $rank(NumberOfClusters)$ based on validity criteria
      Applied OWA $rank(NumberOfClusters)$
      Select results with max($OWA(rank(NumberOfClusters))$)
    else if X-Means **then**
      Run X-Means
    else if DBSCAN **then**
      Run DBSCAN($minPts$,ϵ)
    else if Spectral **then**
      for $NumberOfClusters$=2 **to** 12 **do**
        for σ=0.1:0.3:0.5 **do**
          Run DBSCAN ($NumberOfClusters$,σ)
          Calculate validity criteria
        end for
      end for
      Set $rank(NumberOfClusters, \sigma)$ based on validity criteria
      Applied OWA $rank(NumberOfClusters, \sigma)$
      Select results with max($OWA(rank(NumberOfClusters, \sigma))$)
    end if
    Each Excavation Front EF_j is characterized by the percentage of points in each cluster $C_{ij}(\%)$
  end for

The second step is based on the variable selection previously made: an unsupervised search of categories, or classes, of hole drillings is performed. This step permits descriptions of the different drilling categories embedded into MWD data to be generated in terms of MWD variables. These descriptions can be linked with geological knowledge. Some alternative clustering algorithms and validity criteria have been involved in this goal: K-Means, X-Means, Spectral,etc. The details of the methodology are shown in Algorithm 2. The parametrized clustering approaches generate different performances according to the validity indexes considered, so a new OWA based multicriteria decision making is set out with $orness = \{0.4, 0.5, 0.6\}$ to obtain a global rank, considering the alternative ranks due to the different validity criteria. The distribution of the drillings into categories for each excavation front is used to characterize each of these excavation fronts.

So far, all the tasks are based on unsupervised techniques, but at this point, the collected expert knowledge is incorporated: the evaluations made by the geologist concerning the excavation fronts and the corresponding RMR values. These are used during the supervised learning of the next step devoted to the RMR prediction.

Algorithm 3 RMR Prediction based on Characterization of the Excavation Fronts

```
Require: Excavation Front Characterization from Algorithm 2
  for ModelingAlgorithm=S-IRL:L-IRL do
    for ClusterAlgorithm=K-Means:X-Means:DBSCAN:Spectral:FCM do
      Obtain ExcavationFrontCharacterization EF_j from ClusterAlgorithm
      for each RMR index IndexRMR do
        for InputOption=C_ij(%):C_ij(%) + IndexRMR(EF_{j-1}) do
          for nLT=3:5:7 number of Linguistic Terms for S-IRL and L-IRL do
            for CrossValidation=CV_1:CV_5 do
              Generate randomly 5-fold cross-validation
              Tune ModelingAlgorithm (ClusterAlgorithm,IndexRMR,InputOption,nLT,CV_i)
              Calculate MRE(CV_i) and nR(CV_i)for training and test
            end for
            Calculate average of MRE and nR for training and test
          end for
        end for
      end for
    end for
  end for
  Set ranks based on MRE and nR for test
  Applied OWA rank(MRE, nR, nlT)
  Select prediction scheme with max(OWA(rank(MRE, nR, nlT)))
```

The final step concerns the prediction of RMR values based on the MWD featuring of the excavation fronts. This MWD based forecast for RMR values is made by two versions of genetic FRBSs implementing an approximate and a linguistic approach: S-IRL and L-IRL. This FRBS based approach also permits an extraction of knowledge by fuzzy rules using MWD linguistic terms, linking the MWD knowledge with geological knowledge. In order to tune these fuzzy approaches, some ranges of parameters and criteria were considered to evaluate the best model: accuracy (MRE), number of rules(nR) or number of linguistic terms (nlT). Once more, an OWA based multicriteria decision making is set out for this estimation, considering different values for $orness$ according to the needs of the user. Algorithm 3 describes this stage in detail.

5.2. Results and Analysis

5.2.1. MWD Data Pre-Processing

The data sets available are the MWD recordings of the drillings made by the jumbo machine (see Table 2). A data cleaning is needed to remove and fix such anomalies as drillings with only one record, duplicate drillings on HD or T, etc. Finally, 3551 hole drillings containing 15548 samples were preserved (Fig. 1 (c))

Table 2: MWD Variables

MWD Variable		Measure unit → normalization if needed
HD	Hole Depth	*millimeters → meters*
PR	Penetration Rate	*decimeters per minute → meters per second*
HP	Hammer Pressure	*bar*
FP	Feed Pressure	*bar*
DP	Damper Pressure	*bar*
RS	Rotation Speed	*r per minute → r per second*
RP	Rotation Pressure	*bar*
WF	Water Flow	*litres per minute → litres per second*
WP	Water Pressure	*bar*
T	Time	*hour:minute:second → seconds since 1th January 1970 1:00am*

According to the methodology (see Methodology 1), the Fourier Transform on time and depth (length) is calculated for each drilling described by 8 MWD variables. The FT for each of these MWD variables is summarized by 3 values: A_0, B_W, R_A. A first filtering is made by the Pearson correlation coefficients and some variables are excluded. Then, 30 of the variables are preserved for the followings steps (Table 3).

5.2.2. Unsupervised Feature Selection from MWD data

The target is to detect and select the most relevant MWD variables previously described in an unsupervised way. In Algorithm 1, the details for this goal are described. This methodology is applied over the 30 remaining variables from the first filtering. Table 3 shows the evaluation of each variable through the 4 evaluation criteria applied: First Principal Component, Principal Components over the averaged value of the principal components set, Principal Components containing at least 90% of the original information, and the maximum number of Analysis Factors. Obviously, each criterion shows a different evaluation for every variable, but a final ranking considering all these alternatives is needed to make a variable selection. A decision making is carried out based on the OWA operator and the RIM quantifiers.

MWD Transformed Variable	**PCA Evaluation**			**FA Evaluation**	**OWA Global Rank**
	First PC	PC's Over Median (12)	PC's $>$ 90% variability (13)	Factors (9)	
FP-$A_{0_{HD}}$	0.064	0.041	0.038	0.045	0.0438
DP-$A_{0_{HD}}$	0.043	0.044	0.045	0.044	0.0436
RP-$A_{0_{HD}}$	0.046	0.043	0.04	0.046	0.0431
PR-$A_{0_{HD}}$	0.073	0.03	0.038	0.045	0.0413
WP-$A_{0_{HD}}$	0.041	0.04	0.039	0.044	0.0404
HP-$A_{0_{HD}}$	0.071	0.037	0.035	0.036	0.0403
WF-$A_{0_{HD}}$	0.027	0.043	0.041	0.046	0.0373
RS-$A_{0_{HD}}$	0.031	0.032	0.035	0.043	0.0338
RS-$R_{A_{HD}}$	0.078	0.027	0.026	0.026	0.0327
HP-$R_{A_{HD}}$	0.077	0.027	0.028	0.024	0.0325
PR-$R_{A_{HD}}$	0.083	0.022	0.023	0.032	0.0321
WF-A_{0_T}	0.011	0.04	0.037	0.044	0.0294
FP-$R_{A_{HD}}$	0.083	0.017	0.02	0.029	0.029
PR-$B_{W_{HD}}$	0.03	0.052	0.051	0.003	0.0282
RP-$R_{A_{HD}}$	0.079	0.017	0.022	0.025	0.0282
WP-$R_{A_{HD}}$	0.074	0.018	0.023	0.023	0.0277
DP-A_{0_T}	0.001	0.039	0.037	0.046	0.0261
RS-A_{0_T}	0.01	0.033	0.03	0.046	0.0257
WP-A_{0_T}	0.0	0.039	0.037	0.046	0.0256
RP-A_{0_T}	0.003	0.033	0.034	0.045	0.0239
HP-$B_{W_{HD}}$	0.015	0.052	0.052	0.001	0.023
FP-A_{0_T}	0.002	0.03	0.033	0.046	0.0227
RP-B_{W_T}	0.011	0.029	0.027	0.03	0.0224
PR-A_{0_T}	0.0	0.03	0.031	0.046	0.0219
WP-B_{W_T}	0.01	0.025	0.026	0.032	0.0208
PR-B_{W_T}	0.01	0.023	0.024	0.039	0.0205
PR-R_{A_T}	0.006	0.026	0.025	0.041	0.0204
RS-B_{W_T}	0.009	0.028	0.026	0.023	0.0193
WF-B_{W_T}	0.006	0.042	0.039	0.007	0.0181
RS-$B_{W_{HD}}$	0.007	0.04	0.038	0.0	0.0157

Table 3: MWD transformed variables: evaluations by Principal Components and Analysis Factors and OWA based global ranking ($orness = 0.4$)

Several values for the *orness* parameters have been checked at a round 0.5 without serious changes. Finally, the results shown are based on $orness = 0.4$, because as it is slightly less conservative to deal with the variable selections. Observing these results, many relevant aspects can be noted: the 11 best evaluated variables are obtained by the FT based on the depth/length of drilling, and corresponding to the Amplitude $w = 0$ for different MWD variables. So, the challenge is the number of variables to be considered; a first set of selected variables can correspond to the 6 best evaluated, but this number of variables was checked in the clustering task of this proposal: the results were poor, and worse (even, in some clustering algorithms not enough results were obtained) in comparison with those obtained by a more reduced number of variables, such as the case of the three first variables in the ranking. All this is due to the previously commented complexity/dimensionality issue concerning the clustering. On the other hand, the description in linguistic terms is harder, so the rest of the results are based on the 3 first variables obtained by the MWD Fourier transform based variables:

1. FP-$A_{0_{HD}}$: Feed Pressure Gain.
2. DP-$A_{0_{HD}}$: Damper Pressure Gain.
3. RP-$A_{0_{HD}}$: Rotation Pressure Gain.

5.2.3. Clustering based characterization for excavation fronts

The results of the clustering based on the previous variable selection are shown in Table 4, according to Algorithm 2. One of the challenges for most of the clustering algorithms is the number of clusters: In Tables 5, 6 and 7 the performance is shown when different numbers of clusters and alternative validity indexes are considered for K-Means, Spectral and FCM Clustering, respectively. Once more, a decision making is set out in order to obtain a global ranking for every algorithm, and then to select the best number of clusters for each case, which means data partitioning. On the other hand, X-Means is able to estimate, by itself, the optimum number of clusters for data, and the number of clusters for the DBSCAN case is based on a heuristic for MinPts parameters (Fig. 3) [47]. The results for Spectral Clustering are based on $\sigma = 0.1$. The range of clusters to be checked by the algorithms was based on the expertise concerning the real excavation fronts and the need for experimental work.

The different nature of every clustering algorithm has implied different performances, such as number or shape of clusters. In general, for most of

Clustering	Number of Cluster (nC)	Prototypes [FP-$A_{0_{HD}}$ DP-$A_{0_{HD}}$ RP-$A_{0_{HD}}$]	Number of drillings
K-Means	1	[0.636 0.613 0.574]	1526 (42.97 %)
	2	[0.871 0.704 0.936]	1482 (41.73 %)
	3	[0.815 1.312 0.612]	543 (15.29 %)
X-Means	1	[0.751 0.583 0.67]	2953 (83.16 %)
	2	[0.969 1.286 0.651]	598 (16.84 %)
DBSCAN	noise	[0.0 0.0 0.0]	174 (4.9 %)
	1	[0.762 0.668 0.74]	2949 (83.05 %)
	2	[0.37 0.693 0.807]	6 (0.17 %)
	3	[0.607 1.059 0.642]	7 (0.2 %)
	4	[0.879 1.116 1.244]	5 (0.14 %)
	5	[0.192 0.367 0.46]	20 (0.56 %)
	6	[1.054 1.686 0.63]	67 (1.89 %)
	7	[0.799 1.291 0.555]	278 (7.83 %)
	8	[0.261 1.173 0.301]	17 (0.48 %)
	9	[1.05 1.714 1.033]	15 (0.42 %)
	10	[0.695 1.137 0.728]	5 (0.14 %)
	11	[1.041 1.716 1.309]	5 (0.14 %)
	12	[0.605 0.747 0.812]	3 (0.08 %)
Spectral	1	[1.011 1.692 0.750]	93 (2.62 %)
	2	[0.768 1.251 0.564]	398 (11.21 %)
	3	[0.752 0.665 0.752]	3060 (86.17 %)
FCM	1	[0.645 0.603 0.585]	1475 (41.54 %)
	2	[0.854 0.696 0.905]	1502 (42.3 %)
	3	[0.821 1.267 0.613]	574 (16.16 %)

Table 4: Clustering results according to validity criteria and the OWA based decision making.

nC	CH	r	DB	r	D	r	SD	r	S	r	XB	r	OWA(r)
2	1646	0	1.30	0	0.005	4	22.10	2	0.31	5	1261	4	1.9
3	**2556**	**10**	**0.74**	**10**	0.007	9	14.58	8	**0.38**	**10**	676	8	**8.9**
4	2167	7	0.85	7	0.005	5	**14.19**	**10**	0.34	9	1347	2	5.8
5	2271	9	0.93	4	0.002	0	14.21	9	0.33	8	4163	0	3.8
6	2224	8	1.05	2	**0.009**	**10**	15.64	7	0.32	6	**407**	**10**	6.3
7	1872	1	0.90	5	0.005	3	20.67	3	0.26	0	829	5	2.2
8	2006	6	0.75	9	0.006	7	19.16	5	0.29	1	680	7	5.1
9	1965	4	0.79	8	0.004	2	19.03	6	0.30	3	1314	3	3.6
10	1968	5	0.89	6	0.006	6	19.95	4	0.32	7	687	6	5.4
11	1931	3	1.02	3	0.006	8	22.17	1	0.30	2	560	9	3.3
12	1887	2	1.06	1	0.004	1	22.87	0	0.31	4	1673	1	1.1

Table 5: OWA decision making on validity criteria for K-Means Clustering, $orness = 0.4$

nC	S	r	CH	r	DB	r	D	r	SD	r	XB	r	OWA(r)
2	0.58	9	5061	9	0.68	8	0.007	9	197	8	308	9	8.5
3	**0.76**	**10**	**13122**	**10**	**0.30**	**10**	**0.016**	**10**	**70**	**10**	**55**	**10**	**10.0**
4	0.49	7	1976	7	0.84	5	0.000	0	477	1	341334	0	2.3
5	0.51	8	2086	8	0.75	7	0.000	1	636	0	27842	2	3.3
6	0.40	2	874	6	1.24	4	0.001	4	409	2	7617	5	3.3
7	0.40	3	827	5	1.37	1	0.000	2	304	3	29310	1	2.0
8	0.40	4	732	2	1.28	2	0.001	5	199	7	8688	4	3.4
9	0.41	5	521	0	0.49	9	0.002	8	219	6	2117	8	5.1
10	0.39	0	739	3	1.38	0	0.001	3	288	4	15066	3	1.7
11	0.41	6	739	4	1.28	3	0.001	6	254	5	7425	6	4.7
12	0.39	1	631	1	0.76	6	0.002	7	193	9	3425	7	4.2

Table 6: OWA based decision making on validity criteria for Spectral Clustering, $orness = 0.4$, $\sigma = 0.1$

nC	XB	r	FS	r	PC	r	PE	r	OWA(r)
2	1.03E-4	7	-65.63	1	**0.71**	**10**	**0.45**	**10**	6.0
3	8.1E-5	9	**-105.91**	**10**	0.63	9	0.65	9	**9.1**
4	1.0E-4	8	-96.25	9	0.53	8	0.87	8	8.1
5	**7.1E-5**	**10**	-94.47	8	0.49	7	1.01	7	7.6
6	1.3E-4	5	-89.99	4	0.46	6	1.12	6	5.0
7	1.58E-4	3	-83.23	3	0.42	5	1.25	5	3.7
8	1.26E-4	6	-90.57	5	0.41	4	1.30	4	4.5
9	3.02E-4	0	-76.60	2	0.36	2	1.45	3	1.4
10	1.53E-4	4	-93.61	7	0.37	3	1.46	2	3.4
11	1.68E-4	2	-62.42	0	0.32	0	1.65	0	0.2
12	1.75E-4	1	-91.07	6	0.34	1	1.59	1	1.6

Table 7: OWA based decision making on validity criteria for FCM Clustering, $orness = 0.4$

the algorithms, the number of clusters has been 3, except for X-Means that generates 2, and DBSCAN has detected 12 clusters and a number of drillings, around 5%, are not considered for any cluster (see Table 4). On the other hand, each cluster means a class of drilling whose prototypes are shown in Table 4, described by the MWD transformed variables and the distribution in % of the hole drillings in every category.

Now, each excavation front is ready to be characterized through its own drilling category distribution, according to each clustering algorithm considered: each hole drilling is in a cluster or category. In Table 8, a couple of samples of this featuring by type of drillings of some excavation fronts are shown. Figure 4 shows a graphic representation of a couple of excava-

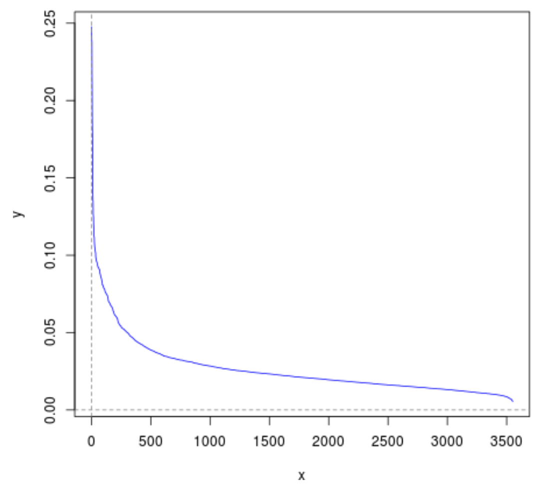


Figure 3: Expert decision making for DBSCAN based on k-distance graph, $k = minPts = 4$, $\epsilon = 0.05$

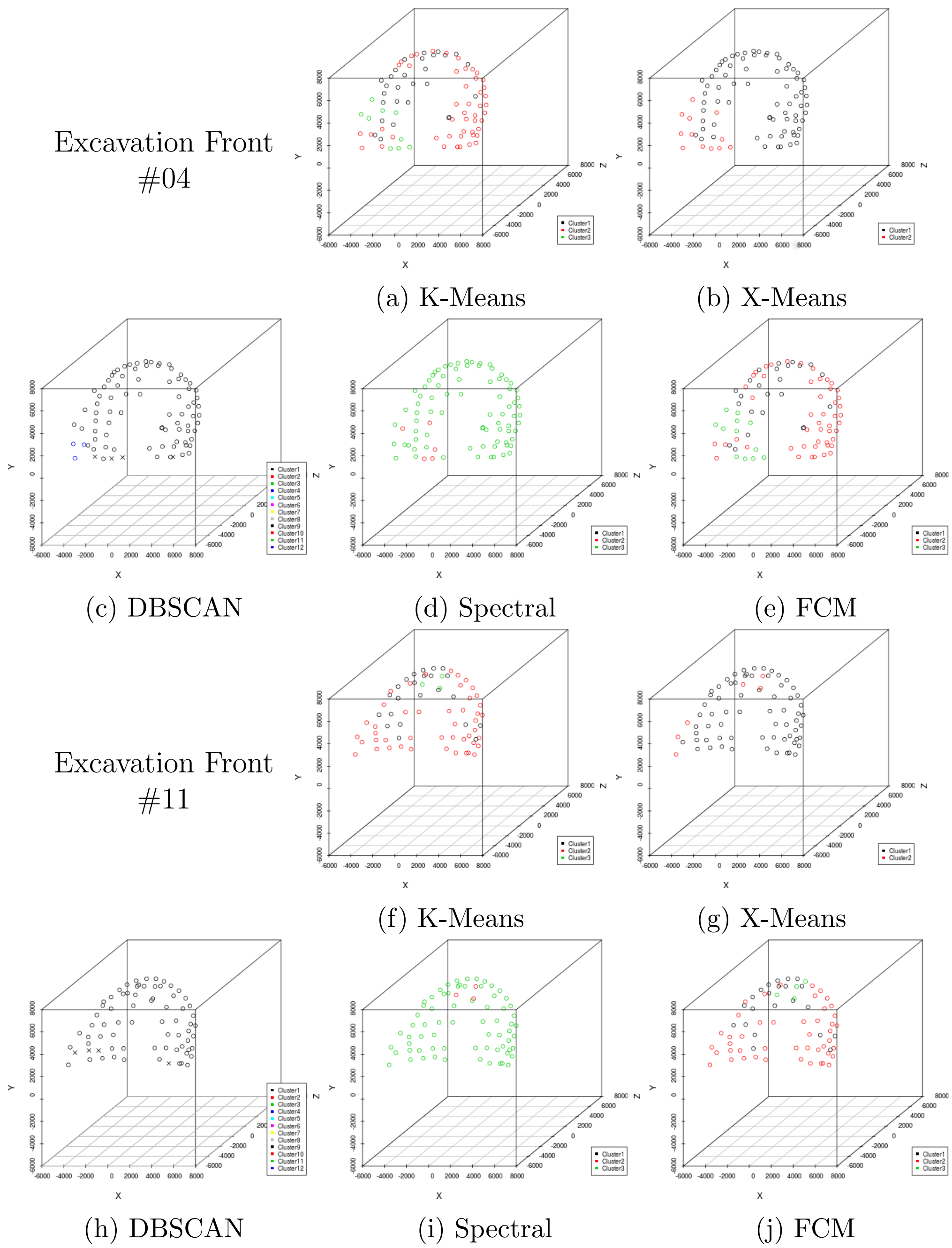
Excavation Front	**X-Means** $C_1(\%)$	$C_2(\%)$	**Spectral** $C_1(\%)$	$C_2(\%)$	$C_3(\%)$
#01	75.61	24.39	0.0	12.2	87.8
#02	87.36	12.64	0.0	8.05	91.95
#03	91.86	8.14	0.0	1.16	98.84
#04	85.88	14.12	0.0	5.88	94.12
#05	85.37	14.63	0.0	1.22	98.78
#06	76.40	23.60	0.0	14.61	85.39
#07	100.0	0.0	0.0	0.0	100.0
#08	40.00	60.00	53.33	0.0	46.67
#09	51.69	48.31	10.11	35.96	53.93
#10	69.23	30.77	0.0	15.38	84.62
#11	91.67	8.33	0.0	4.17	95.83
#12	60.00	40.00	0.0	10.0	90.0

Table 8: Samples of excavation fronts featured by X-Means and Spectral clusters.

tion fronts according to each clustering algorithm used. This featuring is the foundation input for the next stage concerning the RMR estimation of values and parameters.

5.2.4. MWD Excavation Front based RMR Prediction

This stage involves the RMR estimations for 52 tunnel excavation fronts, 24 and 28 for each direction respectively. This prediction is based on the previous characterization of every excavation front by drilling categories and



(a) K-Means (b) X-Means (c) DBSCAN (d) Spectral (e) FCM (f) K-Means (g) X-Means (h) DBSCAN (i) Spectral (j) FCM

Figure 4: Samples of excavation fronts featured by types of drilling, or clusters, for every clustering algorithm considered.

it concerns RMR_{basic} and RMR values.

The RMR forecasting is based on two well-known genetic FRBS approaches (S-IRL and L-IRL), using a scheme of 5-fold cross validation, taking into account, as performance, the prediction error, the number of rules (complexity) and the number of linguistic terms according to Algorithm 3. These algorithms need to define the fuzzy partition for the variables involved: this is a tuning parameter called *number of linguistic terms* (nLT). The parameter values considered are 3, 5 and 7 because they are considered the most recommendable when fuzzy rules with a linguistic and interpretable performance are wanted (see the database in Figure 5). The results shown are the number of rules (nR) and the error for training (MRE_{tra}) and testing (MRE_{tst}) in two cases: when the input is the excavation front featured by its drilling distribution ($C_i(\%)$), or adding the RMR value of the previous excavation front ($RMR(EF_{j-1})$) as a new input. These results concern each clustering algorithm taken into account in the previous clustering stage. Then, a new decision making is carried out, taking into account the number of rules, the test error and the number of linguistic terms used.

Tables 9 and 10 show average results concerning the RMR_{basic} and RMR estimations [1]. The best results for S-IRL and L-IRL are obtained when their inputs are the distribution of drilling categories and the RMR value of the previous excavation front $RMR(EF_{j-1})$.

On the other hand, the prediction based on the linguistic approach (L-IRL) is a bit better for both cases: RMR_{basic} and RMR. The model based on L-IRL and DBSCAN has the lower test error, $MRE_{tst} = 2.85\%$, but the number of rules is very high $nR = 83$ with $nLT = 7$ linguistic terms, so its complexity is higher than the prediction set out by L-IRL and X-Means, which presents a slightly higher test error, $MRE_{tst} = 3.01\%$, but the number of rules is $nR = 17$ with $nLT = 3$ linguistic terms. The decision making based on the RIM quantifier and the OWA, considering the accuracy and complexity of the model, shows the best solution (selecting $orness = 0.4$).

A similar analysis can be made for RMR: L-IRL shows the lowest test error, $MRE_{tst}3.36\%$ with $nr = 58.4$ rules and $nLT = 5$ linguistic terms for

[1] Parameters to run IRL algorithms are: minimum covering degree= 1.5, covering for positive examples= 0.05, negative examples= 0.1%, population size= 61, generations= 100, crossover a= 0.35 with probability 0.6, mutation b= 5 with probability= 0.1, evolutionary strategy applied until there is no improvement in 50 generations over 20% of individuals of population.

Clustering		L-IRL Prediction Inputs					
		$C_i(\%)$ and $RMR_{basic}(EF_{j-1})$			$C_i(\%)$		
	nLT	nR	MRE_{tra}	MRE_{tst}	nR	MRE_{tra}	MRE_{tst}
K-Means	3	24.0	3.32%	4.38%	19.6	4.70%	4.98%
	5	56.0	2.19%	3.73%	50.6	3.68%	5.14%
	7	83.8	1.88%	4.02%	71.0	3.03%	4.34%
X-Means	3	17.0	2.88%	3.01%	12.8	4.17%	4.32%
	5	34.2	2.53%	3.16%	22.8	3.68%	3.81%
	7	50.4	2.28%	3.44%	34.2	3.67%	3.94%
DBSCAN	3	50.2	2.98%	4.11%	48.2	4.74%	5.65%
	5	75.2	2.09%	3.80%	67.4	3.67%	5.11%
	7	83.0	1.57%	2.85%	72.2	3.35%	5.05%
Spectral	3	19.8	3.11%	3.77%	16.8	4.91%	5.65%
	5	37.6	2.71%	3.22%	28.0	4.28%	5.12%
	7	58.0	2.32%	3.90%	42.6	4.02%	5.06%
FCM	3	26.0	3.32%	3.80%	20.8	5.05%	5.24%
	5	59.6	2.35%	4.10%	52.2	3.49%	4.67%
	7	85.4	1.93%	3.95%	72.8	3.01%	4.20%
Clustering		**S-IRL Prediction Inputs**					
		$C_i(\%)$ and $RMR_{basic}(EF_{j-1})$			$C_i(\%)$		
	nLT	nR	MRE_{tra}	MRE_{tst}	nR	MRE_{tra}	MRE_{tst}
K-Means	3	12.2	3.90%	4.43%	12.0	4.33%	4.99%
	5	34.6	2.07%	3.32%	31.4	3.13%	4.77%
	7	55.8	1.32%	3.95%	49.8	2.14%	5.14%
X-Means	3	10.2	3.72%	4.11%	8.4	4.35%	4.51%
	5	21.4	2.56%	3.00%	16.0	3.90%	3.76%
	7	35.2	1.84%	3.10%	24.2	3.33%	3.87%
DBSCAN	3	28.6	3.23%	4.04%	27.2	4.07%	5.26%
	5	45.4	1.86%	4.01%	45.4	3.30%	5.36%
	7	55.4	1.30%	3.24%	53.0	2.91%	5.78%
Spectral	3	11.6	3.86%	4.24%	11.0	4.66%	5.13%
	5	21.2	2.63%	3.45%	19.0	4.24%	5.26%
	7	36.6	1.93%	3.83%	28.4	3.86%	5.20%
FCM	3	11.6	4.00%	4.20%	11.4	4.32%	4.60%
	5	36.4	2.18%	4.02%	33.2	2.71%	4.56%
	7	55.2	1.41%	3.84%	49.0	1.83%	4.31%

Table 9: IRL based schemes for RMR_{basic} index: number of linguistic terms (nLT), number of rules (nR), error (MRE).

K-Means; this complexity is higher than the results obtained by L-IRL for X-Means test error, $MRE_{tst} = 3.39\%$ with $nR = 17.4$ rules and $nLT = 3$ linguistic terms. These latter are slightly worse for the prediction error, but the complexity is lower, which is an important issue when the knowledge base

Clustering	nLT	**L-IRL Prediction Inputs**					
		$C_i(\%)$ and $RMR(EF_{j-1})$			$C_i(\%)$		
		nR	MRE_{tra}	MRE_{tst}	nR	MRE_{tra}	MRE_{tst}
K-Means	3	26.0	3.27%	4.49%	20.6	4.48%	4.52%
	5	58.4	2.26%	3.36%	50.2	3.72%	4.25%
	7	92.2	1.96%	4.37%	75.8	3.39%	4.48%
X-Means	3	17.4	2.92%	3.39%	13.0	4.56%	4.75%
	5	36.8	2.96%	3.64%	21.6	4.09%	4.19%
	7	58.8	2.52%	4.48%	35.8	3.96%	4.42%
DBSCAN	3	52.2	3.44%	4.50%	48.4	4.58%	5.43%
	5	79.8	2.21%	3.84%	68.8	3.68%	5.70%
	7	89.0	1.62%	3.89%	74.4	3.34%	5.45%
Spectral	3	21.0	3.18%	3.41%	17.4	4.84%	5.22%
	5	40.0	2.79%	3.53%	27.2	4.56%	4.97%
	7	61.6	2.50%	3.97%	42.4	4.24%	4.58%
FCM	3	28.0	3.21%	4.07%	22.0	4.44%	4.67%
	5	62.4	2.37%	4.03%	53.4	3.62%	4.42%
	7	92.8	2.02%	4.66%	76.0	3.29%	5.04%
Clustering	nLT	**S-IRL Prediction Inputs**					
		$C_i(\%)$ and $RMR(EF_{j-1})$			$C_i(\%)$		
		nR	MRE_{tra}	MRE_{tst}	nR	MRE_{tra}	MRE_{tst}
K-Means	3	12.8	3.74%	4.30%	12.2	4.32%	4.72%
	5	35.2	2.15%	4.03%	32.0	3.19%	4.22%
	7	56.4	1.52%	4.09%	53.6	2.49%	4.58%
X-Means	3	9.8	3.55%	4.14%	9.0	4.47%	4.50%
	5	21.4	2.94%	3.50%	16.6	3.88%	4.06%
	7	39.2	2.04%	4.11%	24.4	3.50%	3.97%
DBSCAN	3	30.8	3.52%	5.01%	27.0	4.04%	5.11%
	5	47.6	1.85%	4.18%	46.6	3.28%	5.62%
	7	59.8	1.36%	3.89%	56.8	2.94%	5.49%
Spectral	3	11.6	3.76%	4.55%	11.4	4.76%	5.26%
	5	22.6	2.74%	3.57%	18.0	4.35%	4.81%
	7	40.0	2.09%	4.06%	29.4	3.92%	4.41%
FCM	3	12.6	3.69%	4.34%	11.4	4.26%	4.62%
	5	35.2	2.12%	4.43%	33.0	2.98%	4.55%
	7	58.8	1.51%	4.53%	52.6	2.20%	4.98%

Table 10: IRL based schemes for RMR index: number of linguistic terms (nLT), number of rules (nR), error (MRE)

by fuzzy rules is a desired target. Contradictory criteria would once more be solved by an OWA based decision making according to the user needs or preferences.

5.3. Extraction of Knowledge by Fuzzy rules: A sample

The prediction based on FRBSs, such as L-IRL and S-IRL, introduces data driven learning, which allows a knowledge base, expressing and explaining this knowledge by "interpretable" fuzzy rules, to be obtained. Figure 5 shows a knowledge base sample generated by L-IRL using X-Means clustering to characterize the excavation front, with 3 linguistic variables (categories of hole drillings) as rule antecedents and 1 linguistic variable (RMR value) as rule consequent, using 3 linguistic terms ($nLT = 3$) for each one. This knowledge extraction implies that the number of rules, as well as the number of linguistic terms or the number of rule antecedents, must be taken into account during the selection of the prediction model. A couple of examples of these rules are:

R1: IF $\%Type1Drillings$ is High AND $\%Type2Drillings$ is Low
AND $PreviousRMR$ is Medium
THEN RMR is Medium
R15: IF $\%Type1Drillings$ s Medium AND $\%Type2Drillings$ is Medium
AND $PreviousRMR$ is Medium
THEN RMR is High

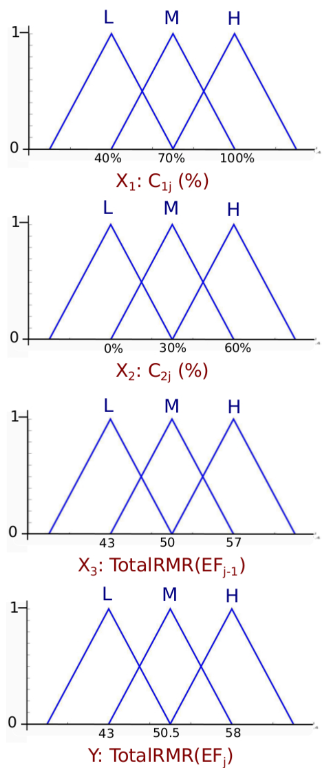
The quality and level of interpretability of this knowledge base are the subject of another research field: these rules are not without problems, such as: redundancy, incoherence, etc. [55], [56], but all this can be managed and a serviceable knowledge base can be achieved that, in another way, would not be available, connecting the unsupervised MWD information level with the expert RMR information level. The compilation of these knowledge bases concerning several cases will permit a very serviceable library to be created for giving support *on-site* and in real time.

5.4. Final comments: Summary of results

The open approach introduced in this paper is able to manage, on-site, the MWD data generated by the drill rig to estimate the critical RMR value of every excavation front of the tunnel in progress. The analysis of the MWD data, on the time and frequency domain, has permitted only 3 main MWD features to be used as the basis for the rest of the methodology, meaning a huge reduction in the complexity of the solution.

This reduced number permits the drillings for a clustering procedure to be featured. In this way, every excavation front is summarized by a very few features based on these MWD drilling based rocky categories. This can be

DATA BASE



RULE BASE

Rule	$\boldsymbol{X_1}$	$\boldsymbol{X_2}$	$\boldsymbol{X_3}$	$\boldsymbol{Y}$
R_1	H	L	M	M
R_2	H	L	M	M
R_3	H	L	M	M
R_4	M	M	M	M
R_5	M	M	M	M
R_6	L	H	H	M
R_7	L	H	H	M
R_8	L	H	H	M
R_9	H	L	L	L
R_{10}	H	L	L	L
R_{11}	H	L	L	L
R_{12}	M	M	H	H
R_{13}	H	L	H	H
R_{14}	H	L	H	H
R_{15}	M	M	M	H

Variables

X_1:$C_{1j}(\%)$. Type 1 drillings percentage
X_2:$C_{2j}(\%)$. Type 2 drillings percentage
X_3:$RMR(EF_{j-1})$. Previous EF index value
Y:$RMR(EF_j)$. Index value to predict

Linguistic Terms

L: Low, M: Medium, H: High

Accuracy

$MRE_{tra} = 3.32\%$; $MRE_{tst} = 3.08\%$

Figure 5: Knowledge Base for RMR: L-IRL, X-Means, $nLT = 3$

seen as a feature extraction that summarizes the characterization of every excavation front to an affordable dimensionality for ML & CI approaches. These reductions of dimensionality/complexity are critical for addressing this challenge.

Most different cluster policies, or algorithms, have shown that 2 or 3 clusters is a well-balanced number of MWD drilling based rocky categories. This fits with the knowledge and expertise concerning this issue of the technicians in charge of this type of work. This characterization has permitted the RMR value to be estimated using a linguistic and scatter FRBS: permitting the capability of both different approaches to be checked, so as to estimate the RMR while also generating a reasonable base of well-balanced fuzzy rules regarding accuracy-interpretability. This means being able to generate a good estimation and an "interpretable" knowledge base about the drilling features of every excavation front and their RMR values based on linguistic terms. This modelling has been made possible by the expert knowledge provided by geologists concerning the excavation fronts. The linguistic approach (L-IRL) has been slightly better than the scatter option (S-IRL), providing the best approach for RMR estimation as a reasonable base of knowledge: $MRE_{tst} = 3.01\%$, 17 fuzzy rules, 3 linguistic variables and 3 linguistic terms. So the complexity of this knowledge base is affordable. Other more accurate predictions are possible, $MRE_{tst} = 3.01\%$, but with an increase in the complexity, 83 fuzzy rules.

6. Conclusions

This work is focused on an open, well-defined and methodological approach to take advantage of MWD data to be used in the prediction of design parameters, here the RMR, during a tunnelling. The proposal is based on ML & CI techniques, which are able to deal with this challenge, involving such issues as: noisy data, large number of variables, reduction of complexity, fuzzy knowledge, shortage of expert knowledge, prediction models, decision making, etc.

Here, unsupervised and supervised techniques permit issues to be dealt with when the expert knowledge is not available or reliable, and when this knowledge is available to be taken advantage of. The methodology takes into account several top approaches and criteria for every stage and, based on a multidecision making using *linguistic terms*, the users can tune their own risk to be assumable for the prediction. On the other hand, the prediction is based on an FRBS that incorporates explanation capability, even in *linguistic terms*, about the MWD data based prediction in comparison with the usual black-box modelling in the literature. All this permits, with the well-defined methodology, robust and competitive results.

The case study shows an MWD based estimation for the RMR_{basic} and RMR values only made an MRE_{tst} of around 3%, based on FRBS models with an affordable complexity: 17 fuzzy rules, containing 3 linguistic variables as rule antecedents using 3 linguistic terms. This RMR prediction rate can be even better, with for instance an $MRE_{tst} = 2.85\%$, but increasing the complexity of the knowledge base; in any case, this is an user decision. Once again, the accuracy, its knowledge base and well-defined FRBS tuning, are very competitive in comparison with other approaches in the literature.

On the other hand, as commented in previous sections, the reductions of dimensionality/complexity carried out with the MWD data and the characterization of the excavation fronts are critical issues for this type of challenge. The methodology defined in this work is general enough to be applicable to other types of tunnels and excavations, cases and parameters.

References

[1] W. G. C. T. ITA, General report on conventional tunnelling method, Tech. Rep. 002, International Tunnelling and Underground Space Association (2009). doi:tunnel.ita-aites.org/media/k2/attachments/public/ITA$_{R}eport_{N}2_{W}G19_{P}.pdf$.

[2] P. Rai, H. Schunnesson, P.-A. Lindqvist, U. Kumar, An overview on measurement-while-drilling technique and its scope in excavation industry, Journal of The Institution of Engineers (India): Series D 96 (1) (2015) 57–66. doi:10.1007/s40033-014-0054-4.

[3] Z. T. Bieniawski, Engineering classification of jointed rock masses, Transaction of the South African Institution of Civil Engineers 15 (12) (1973) 335–443.
URL `http://journals.co.za/content/civeng/15/12/AJA10212019`$_1$7397

[4] H. Jang, E. Topal, A review of soft computing technology applications in several mining problems, Applied Soft Computing 22 (2014) 638 – 651. doi:http://dx.doi.org/10.1016/j.asoc.2014.05.019.

[5] A. Garaga, G. M. Latha, Intelligent prediction of the stressstrain response of intact and jointed rocks, Computers and Geotechnics 37 (5) (2010) 629 – 637. doi:http://dx.doi.org/10.1016/j.compgeo.2010.04.001.

[6] S. K. Kashyap, D. Parhi, A. Sinha, M. Singh, B. Singh, Optimization of mine support parameters using neural network approach, in: Proceedings of he 12th International Conference of International Association for Computer Methods and Advances in Geomechanics (IACMAG), October 1–6, 2008, Goa, India, 2008, pp. 1170–1179.
URL `http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.383.9178rep=rep1type`

[7] Z. Zafirovski, M. Jovanovski, D. Moslavac, Z. Krakutovski, Application of neural networks in analyzing of rock mass parameters in tunnelling.,

in: Proceedings of the 3^{thd} International Conferences on Road and Rail Infrastructure (CETRA2014). April 28–20, 2014, Split, Croatia, 2014.
URL http://master.grad.hr/cetra/ocs/index.php/cetra3/cetra2014/paper/view/331

[8] B. R. Kumar, H. Vardhan, M. Govindaraj, G. Vijay, Regression analysis and ANN models to predict rock properties from sound levels produced during drilling, International Journal of Rock Mechanics and Mining Sciences 58 (2013) 61 – 72. doi:http://dx.doi.org/10.1016/j.ijrmms.2012.10.002.

[9] S.-S. Leu, C.-N. Chen, S.-L. Chang, Data mining for tunnel support stability: neural network approach, Automation in Construction 10 (4) (2001) 429 – 441. doi:http://dx.doi.org/10.1016/S0926-5805(00)00078-9.

[10] S. Mahdevari, S. R. Torabi, M. Monjezi, Application of artificial intelligence algorithms in predicting tunnel convergence to avoid {TBM} jamming phenomenon, International Journal of Rock Mechanics and Mining Sciences 55 (2012) 33 – 44. doi:http://dx.doi.org/10.1016/j.ijrmms.2012.06.005.

[11] S.-S. Leu, T. J. W. Adi, Probabilistic prediction of tunnel geology using a hybrid neural-hmm, Engineering Applications of Artificial Intelligence 24 (4) (2011) 658 – 665. doi:http://dx.doi.org/10.1016/j.engappai.2011.02.010.

[12] H. Sonmez, C. Gokceoglu, R. Ulusay, An application of fuzzy sets to the geological strength index (GSI) system used in rock engineering, Engineering Applications of Artificial Intelligence 16 (3) (2003) 251 – 269. doi:http://dx.doi.org/10.1016/S0952-1976(03)00002-2.

[13] A. C. Adoko, C. Gokceoglu, L. Wu, Q. J. Zuo, Knowledge-based and data-driven fuzzy modeling for rockburst prediction, International Journal of Rock Mechanics and Mining Sciences 61 (2013) 86 – 95. doi:http://dx.doi.org/10.1016/j.ijrmms.2013.02.010.

[14] J. Jonak, Use of artificial intelligence in automation of rock cutting by mining machines, Journal of Mining Science 38 (3) (2002) 270–277. doi:10.1023/A:1021903904172.

[15] M. Esmaeili, A. Salimi, C. Drebenstedt, M. Abbaszadeh, A. Aghajani Bazzazi, Application of PCA, SVR, and ANFIS for modeling of rock fragmentation, Arabian Journal of Geosciences 8 (9) (2014) 6881–6893. doi:10.1007/s12517-014-1677-3.

[16] A. Yazdani-Chamzini, M. Razani, S. H. Yakhchali, E. K. Zavadskas, Z. Turskis, Developing a fuzzy model based on subtractive clustering for road header performance prediction, Automation in Construction 35 (2013) 111 – 120. doi:http://dx.doi.org/10.1016/j.autcon.2013.04.001.

[17] H. Jalalifar, S. Mojedifar, A. Sahebi, H. Nezamabadi-pour, Application of the adaptive neuro-fuzzy inference system for prediction of a rock engineering classification system, Computers and Geotechnics 38 (6) (2011) 783 – 790. doi:http://dx.doi.org/10.1016/j.compgeo.2011.04.005.

[18] H. Jalalifar, S. Mojedifar, A. Sahebi, Prediction of rock mass rating using fuzzy logic and multi-variable RMR regression model, International Journal of Mining Science and Technology 24 (2) (2014) 237 – 244. doi:http://dx.doi.org/10.1016/j.ijmst.2014.01.015.

[19] H. N. Rad, Z. Jalali, H. Jalalifar, Prediction of rock mass rating system based on continuous functions using chaos anfis model, International Journal of Rock Mechanics and Mining Sciences 73 (2015) 1 – 9. doi:http://dx.doi.org/10.1016/j.ijrmms.2014.10.004.

[20] S. Alemdag, Z. Gurocak, A. Cevik, A. Cabalar, C. Gokceoglu, Modeling deformation modulus of a stratified sedimentary rock mass using neural network, fuzzy inference and genetic programming, Engineering Geology 203 (2016) 70 – 82, special Issue on Probabilistic and Soft Computing Methods for Engineering Geology. doi:http://dx.doi.org/10.1016/j.enggeo.2015.12.002.

[21] F. F. Martins, T. F. S. Miranda, Estimation of the rock deformation modulus and rmr based on data mining techniques, Geotechnical and Geological Engineering 30 (4) (2012) 787–801. doi:10.1007/s10706-012-9498-1.

[22] J. K. Hamidi, S. Kourosh, B. Rezai, J. Rostami, Performance prediction of hard rock TBM using rock mass rating (RMR) system, Tunnelling and Underground Space Technology 25 (4) (2010) 333 – 345. doi:http://dx.doi.org/10.1016/j.tust.2010.01.008.

[23] B.-S. Chun, W. R. Ryu, M. Sagong, J.-N. Do, Indirect estimation of the rock deformation modulus based on polynomial and multiple regression analyses of the rmr system, International Journal of Rock Mechanics and Mining Sciences 46 (3) (2009) 649 – 658.

doi:https://doi.org/10.1016/j.ijrmms.2008.10.001.
URL http://www.sciencedirect.com/science/article/pii/S1365160908001500

[24] Z. Diddens, Monitoring while drilling for investigating critical rock and rock mass properties for blast design and underground support, Master's thesis, Applied Earth Sciences, TU-Delft (2016).
URL https://repository.tudelft.nl/islandora/object/uuid%3A26f10d09-dd61-49b8-b27b

[25] A. Aydin, Fuzzy set approaches to classification of rock masses, Engineering Geology 74 (34) (2004) 227 – 245. doi:http://dx.doi.org/10.1016/j.enggeo.2004.03.011.

[26] A. Kadkhodaie-Ilkhchi, S. T. Monteiro, F. Ramos, P. Hatherly, Rock recognition from MWD data: A comparative study of boosting, neural networks, and fuzzy logic, IEEE Geoscience and Remote Sensing Letters 7 (4) (2010) 680–684. doi:10.1109/LGRS.2010.2046312.

[27] J. E. J. Martin Gonzalez, Application of pattern recognition techniques to monitoring-while-drilling on a rotary electric blasthole drill at an open-pit coal mine, Master's thesis, Department of Mining Engineering, Queens University, Ontario, Canada (2007).
URL http://hdl.handle.net/1974/924

[28] D. J. Lary, A. H. Alavi, A. H. Gandomi, A. L. Walker, Machine learning in geosciences and remote sensing, Geoscience Frontiers 7 (1) (2016) 3 – 10, special Issue: Progress of Machine Learning in Geosciences. doi:http://dx.doi.org/10.1016/j.gsf.2015.07.003.

[29] N. C. M. Beattie, Monitoring-while-drilling for open-pit mining in a hard rock environment: an investigation of pattern recognition techniques applied to rock identification, Master's thesis, Department of Mining Engineering, Queens University, Ontario, Canada (2009).
URL https://qspace.library.queensu.ca/handle/1974/1780

[30] H. Zhou, S. T. Monteiro, P. Hatherly, F. Ramos, E. Nettleton, F. Oppolzer, Spectral feature selection for automated rock recognition using gaussian process classification, in: Australasian Conference on Robotics and Automation (ACRA), Sydney, Australia, 2009.
URL http://www.araa.asn.au/acra/acra2009/papers/pap118s1.pdf

[31] H. Zhou, P. Hatherly, S. T. Monteiro, F. Ramos, F. Oppolzer, E. Nettleton, S. Scheding, Automatic rock recognition from drilling performance data, in: 2012 IEEE International Conference on Robotics and Automation, 2012, pp. 3407–3412. doi:10.1109/ICRA.2012.6224745.

[32] R. Leung, S. Scheding, Automated coal seam detection using a modulated specific energy measure in a monitor-while-drilling context, International Journal of Rock Mechanics and Mining Sciences 75 (2015) 196 – 209. doi:http://dx.doi.org/10.1016/j.ijrmms.2014.10.012.

[33] I. T. Jolliffe, Principal Component Analysis, Springer, 2002. doi:10.1007/b98835.

[34] N. M. Al-Kandari, I. T. Jolliffe, Variable selection and interpretation in correlation principal components, Environmetrics 16 (6) (2005) 659–672. doi:10.1002/env.728.

[35] M. A. Kramer, Nonlinear principal component analysis using autoassociative neural networks, AIChE Journal 37 (2) (1991) 233–243. doi:10.1002/aic.690370209.

[36] M. Scholz, Validation of nonlinear PCA, Neural Processing Letters 36 (1) (2012) 21–30. doi:10.1007/s11063-012-9220-6.

[37] K. Z. Mao, Identifying critical variables of principal components for unsupervised feature selection, Part B (Cybernetics) IEEE Transactions on Systems, Man, and Cybernetics 35 (2) (2005) 339–344. doi:10.1109/TSMCB.2004.843269.

[38] W. Krzanowski, F. Marriott, Multivariate Analysis: Kendall's Library of Statistics, Wiley, 1994. ISBN:9780340593257.

[39] J. B. MacQueen, Some methods for classification and analysis of multivariate observations, in: Proceedings of 5-th Berkeley Symposium on Mathematical Statistics and Probability, Vol. 1, University of California Press, Berkeley, 1967, pp. 281 – 297. doi:doi=10.1.1.308.8619.

[40] T. Caliński, J. Harabasz, A dendrite method for cluster analysis, Communications in Statistics-Simulation and Computation 3 (1) (1974) 1–27.

[41] D. L. Davies, D. W. Bouldin, A cluster separation measure, IEEE Transactions on Pattern Analysis and Machine Intelligence PAMI-1 (2) (1979) 224–227. doi:10.1109/TPAMI.1979.4766909.

[42] J. C. Dunn, Well-separated clusters and optimal fuzzy partitions, Journal of Cybernetics 4 (1) (1974) 95–104. doi:10.1080/01969727408546059.

[43] M. Halkidi, M. Vazirgiannis, Y. Batistakis, Quality scheme assessment in the clustering process, in: Principles of Data Mining and Knowledge Discovery (PKDD), Vol. 1910 of the series Lecture Notes in Computer Science, Springer Berlin Heidelberg, Berlin, Heidelberg, 2000, pp. 265–276. doi:https://doi.org/10.1007/3-540-45372-5_26.

[44] P. J. Rousseeuw, Silhouettes: A graphical aid to the interpretation and validation of cluster analysis, Journal of Computational and Applied Mathematics 20 (Supplement C) (1987) 53 – 65. doi:https://doi.org/10.1016/0377-0427(87)90125-7.
URL http://www.sciencedirect.com/science/article/pii/0377042787901257

[45] X. Xie, G. Beni, A validity measure for fuzzy clustering, IEEE Transactions on Pattern Analysis and Machine Intelligence 13 (8) (1991) 814–847. doi:10.1109/34.85677.

[46] D. Pelleg, A. W. Moore, X-means: Extending k-means with efficient estimation of the number of clusters, in: Proceedings of the Seventeenth International Conference on Machine Learning, ICML '00, Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 2000, pp. 727–734.
URL http://dl.acm.org/citation.cfm?id=645529.657808

[47] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, A density-based algorithm for discovering clusters a density-based algorithm for discovering clusters in large spatial databases with noise, in: Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, KDD'96, AAAI Press, 1996, pp. 226–231.
URL http://dl.acm.org/citation.cfm?id=3001460.3001507

[48] A. Y. Ng, M. I. Jordan, Y. Weiss, On spectral clustering: Analysis and an algorithm, in: Advances in Neural Information Processing Systems, MIT Press, 2001, pp. 849–856. doi:0262025507.
URL http://dl.acm.org/citation.cfm?id=2980539.2980649

[49] Y. Fukuyama, M. Sugeno, A new method of choosing the number of clusters for the fuzzy c-mean method, in: Proc. 5th Fuzzy Syst. Symp., 1989, 1989, pp. 247–250.
URL https://ci.nii.ac.jp/naid/10026801314/en/

[50] J. C. Bezdek, Pattern Recognition with Fuzzy Objective Function Algorithms, Plenum Press, 1981. doi:http://dx.doi.org/10.1007/978-1-4757-0450-1.

[51] J. C. Bezdek, Cluster validity with fuzzy sets, Journal of Cybernetics 3 (3) (1973) 58–73. doi:10.1080/01969727308546047.

[52] O. Cordón, F. Herrera, Hybridizing genetic algorithms with sharing scheme and evolution strategies for designing approximate fuzzy rule-based systems, Fuzzy Sets and Systems 118 (2001) 235 – 255. doi:https://doi.org/10.1016/S0165-0114(98)00349-2.

[53] O. Cordón, F. Herrera, A three-stage evolutionary process for learning descriptive and approximate fuzzy-logic-controller knowledge bases from examples, International Journal of Approximate Reasoning 17 (4) (1997) 369 – 407, genetic Fuzzy Systems for Control and Robotics. doi:https://doi.org/10.1016/S0888-613X(96)00133-8.

[54] R. Yager, On ordered weighted averaging aggregation operators in multicriteria decision making, IEEE Transactions on Systems, Man and Cybernetics 18 (1998) 183–190. doi:10.1109/21.87068.

[55] M. Galende-Hernández, G. I. Sainz-Palmero, M. J. Fuente-Aparicio, Complexity reduction and interpretability improvement for fuzzy rule systems based on simple interpretability measures and indices by bi-objective evolutionary rule selection, Soft Computing 16 (3) (2012) 451–470. doi:10.1007/s00500-011-0748-6.

[56] M. I. Rey, M. Galende, M. Fuente, G. I. Sainz, Checking orthogonal transformations and genetic algorithms for selection of fuzzy rules based on interpretability-accuracy concepts, International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems 20 (Suppl. 2) (2012) 159 – 186. doi:10.1142/S0218488512400193.