SPECIFIC GROUNDWATER VULNERABILITY MAPPING: CASE STUDY OF ACID MINE DRAINAGE IN THE WITBANK COALFIELD, SOUTH AFRICA

Emmanuel Sakala1, 2, Francois Fourie2, Modreck Gomo2, Henk Coetzee1, Lwazi Magadaza1
1. Council for Geoscience, P Bag X112, Pretoria, 0001, South Africa
2. University of the Free State, IGS, P Box 339, Bloemfontein, 9300, South Africa
esakala@geoscience.org.za, FourieFD@ufs.ac.za, GomoM@ufs.ac.za, henke@geoscience.org.za, lmagadaza@geoscience.org.za

ABSTRACT
This study highlights the usage of artificial neural networks in assessment of groundwater vulnerability. The network uses the DRIST input parameters (Depth to water level, Recharge, Impact of the vadose zone, Soils and Topography) as inputs and hydrochemistry data (Sulphate and Total Dissolved Solids (TDS)) as training data. The results of training and classification using sulphate and TDS were combined using a fuzzy (AND) operator to generate the groundwater vulnerability model. This technique was applied to Witbank Coalfield where acid mine drainage emanating from coal mining operations is a huge concern for surrounding environment and groundwater resources. The generated groundwater vulnerability model of Witbank Coalfield was validated using pH data from 25 groundwater samples. The results show very high negative correlation (0.7182) between the groundwater vulnerability model and pH. This shows that areas with high sulphate and high TDS values correlate with low pH indicative of the presence of possible acid mine drainage pollution. The approach was able to differentiate areas in terms of vulnerability to acid mine drainage which can aid policy and decision makers to make scientifically informed decisions on land use planning. The approach developed in this research need to be applied to other coalfields in order to evaluate its robustness to different hydrogeological and geological conditions.

KEY WORDS
Artificial neutral network, Watershed Policy and Planning, acid mine drainage, groundwater vulnerability, Geographic Information System

1. Introduction
South Africa has nineteen coal provinces, of which the current mining activities are largely focused on coalfields in Mpumalanga Province of South Africa [2]. The coal mining industry has been a fundamental catalyst in the development of the South African economy for over a century and has contributed to it being the most industrialised country in Southern Africa [5]. However coal mining operations in particular Witbank Coalfield has seriously affected the surrounding environment by massive deterioration of groundwater quality by Acid Mine Drainage (AMD) [5]. No regional coalfield scale assessment has been done to establish the vulnerability of the groundwater with regards to pollution by AMD. The research involves development of a groundwater vulnerability model for Witbank Coalfield which would go a long way in demarcating AMD sensitive areas for appropriate land use, watershed policy and planning.

Assessment of groundwater vulnerability is divided in two major types namely the intrinsic and specific vulnerability [20]. The intrinsic type defines only the ease with which a pollutant or a group of pollutants migrates from surface to groundwater. On the other hand the specific type takes into consideration the pollutant properties and its interaction with the subsurface ([24]; [13]). According to [13] in the modern days the intrinsic vulnerability is considered meaningless as compared to specific vulnerability because factors affecting intrinsic vulnerability such as depth to water table, soil and recharge are changing as effects due to humans are increasing. Hence the research focus on the specific vulnerability assessment in Geographic Information System (GIS) environment taking advantage of the good learning and generalisation capabilities of artificial neural networks to establishing the complex relationship between the groundwater vulnerability inputs and AMD indicators.

2. Method and Material

2.1 Study area
The study area is located between 25°30” and 27°45” south latitude, 28°30” and 30°30” east longitude (Figure 1). The area extends from Daveyton to Wonderfontein in the west-east direction and between Kromdraai and Bethal in the north-south covering an area ~ 8 000 km². The study area is marked to the north by the edge of the Karoo rocks and to the south by a palaeohigh named the Smithfield Ridge [12]. Several abandoned and current mining areas are scattered throughout the coalfield. The Olifants River and its tributaries form the main drainage system within the study area.

DOI: 10.2316/P.2016.836-010
2.2 Geology and Hydrogeology

The study area consists of six major lithologies (Figure 1). The geological descriptions of the rocks are displayed in Table 1. The geology was extracted from the 1:250 000 scale geology. [7] subdivided lithologies into four aquifer types, viz. Intergranular - Fractured aquifers consisting of the Rooiberg Group, Loskop Formation, Bushveld Complex, Ecca Group, Karoo dolerite dykes and sills and the Pretoria Group quartzite.

Table 1. Geological description of rocks in the study area.

<table>
<thead>
<tr>
<th>Lithology group (oldest to youngest)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Witwatersrand Supergroup [15]</td>
<td>Shale and Quartzite</td>
</tr>
<tr>
<td>Hospital Hill Formation</td>
<td></td>
</tr>
<tr>
<td>Transvaal Supergroup [9]</td>
<td></td>
</tr>
<tr>
<td>Chuniespoort Group</td>
<td>Dolomite, Chert</td>
</tr>
<tr>
<td>Pretoria Group</td>
<td>Shale, Quartzite</td>
</tr>
<tr>
<td>Rooiberg Group</td>
<td>Rhyolite and Felsite</td>
</tr>
<tr>
<td>Loskop Formation</td>
<td>Sandstone, Conglomerate</td>
</tr>
<tr>
<td>The Bushveld Igneous Complex [4]</td>
<td></td>
</tr>
<tr>
<td>Rustenburg Suite</td>
<td>Gabbro, Norite, Anorthosite</td>
</tr>
<tr>
<td>Lebowa Suite</td>
<td>Felsic granites</td>
</tr>
<tr>
<td>Waterberg Group [4]</td>
<td>Sandstone with interbedded Conglomerate and shale</td>
</tr>
<tr>
<td>The Karoo Supergroup [12]</td>
<td></td>
</tr>
<tr>
<td>Ecca Group</td>
<td>Sandstones, Shales and Coal</td>
</tr>
<tr>
<td>Dwyka Intrusive rocks</td>
<td>Diamictite (Tillite)</td>
</tr>
<tr>
<td>Quaternary [24]</td>
<td>Dolerite (Dykes and Sills)</td>
</tr>
<tr>
<td>Alluvial</td>
<td>Unconsolidated sediments along rivers</td>
</tr>
</tbody>
</table>

The intergranular aquifers correspond to the quaternary alluvial following major rivers. The dolomite of Chuniespoort Group of the Transvaal Supergroup forms karst aquifers found around Delmas Town. The fractured rock aquifers consists of the Black Reef Formation, Dwyka Group, Magaliesberg Formation, Wilger River Formation and the sedimentary rocks of the Pretoria Group with the bulk belonging to the intergranular - fractured aquifer type. The aquifers within the Witbank Coalfield are generally shallow making them highly vulnerable to pollution [24].

2.3 DRIST method

The overlay and index DRIST method [6] which is considered to be an improvement of the popular DRASTIC method [1] was used in this study. The method is based on five parameters of the DRASTIC method namely: Depth to water level, net Recharge, Impact of the vadose zone, Soil media and Topography. The aquifer media and hydraulic conductivity present in the DRASTIC method are not used in the DRIST method as these parameters are difficult to estimate for fractured aquifers which dominate the study area and that the DRIST method only deals with subsurface conditions before pollutants entering the aquifer [6]. The vulnerability values are calculated in the same way as the DRASTIC method by [1].

2.4 Artificial neural networks (ANN)

An ANN consists of a layer of neurons that accept various inputs (Input layer) then fed them into further layers of neurons (Hidden layers) and ultimately to the output layer, which produces an output response (Figure 2). The aim of the technique is to train the network such that its response to a given set of inputs is as close as possible to a desired output [18]. A number of algorithms are available for training a neural network of which the back propagation is the most popular training algorithm [17] and was used in the present study. The five DRIST parameters were used as training datasets. ANN training, validation and classification was done using the Spatial Data Modeller in ArcGIS.

Figure 1. Geological setting and location of the study area.
Figure 2. ANN architecture used in the study [10].

3. Results

The processes of formulating the five DRIST parameters (depth to water level, recharge, impact of the vadose, soils and topography) and combining them using the traditional DRIST method are highlighted in this section. Also the DRIST results were compared with the ANN model generated using the same parameters.

3.1 DRIST input layers

3.1.1 Depth to water level layer (D)

This refers to the distance between the ground level and the water table which determines the passage through which water and pollutant have to travel to get to the groundwater system [1]. The longer the distance the higher the possibility of nature attenuating the pollutants as compared to shorter distances. The depth to water level was calculated from monitoring borehole data obtained from the South African Department of Water and Sanitation. The water level data within the study area were interpolated using the ArcGIS Spatial Analyst Kriking technique to generate the 30 m x 30 m depth to water level gridded raster layer which was reclassified into three classes as proposed by [1] (Figure 3a). The depth to water level varies between 4 m to 21 m and increases from northwest to southeast direction.

3.1.2 Recharge layer (R)

Net recharge represents the total amount of water that percolates down to the vadose zone and reaches the water table and is defined annually [11]. The net recharge parameter caters for the mode of transportation of leached solids or liquids to reach the groundwater system [19]. From a Hydrogeologist point of view, the greater the recharge, the greater the potential for groundwater aquifer pollution. The net recharge used in this study was calculated based on the method devised by [22]. The method involves assigning unit-less values based on the ability to increase the potential recharge value. The parameters include rainfall amount, percentage rise in slope and soil permeability which are combined using the equation (1) [22]:

\[
\text{Recharge} = S + SP + R
\]

Where S – Slope, SP – Soil Permeability and R - Rainfall

Soil permeability values were estimated from soil data acquired from Agricultural Research Council Institute for Soil, Climate and Water (ARC-ISCW). Rainfall data was extracted the country-wide annual rainfall data from the South African Weather Services and slope data was calculated from the Shuttle Radar Topography Mission 30 m resolution data. ArcGIS maths tool was used to combine the weighted factors (Table 2) for slope, rainfall and soil permeability to produce the recharge layer (Figure 3b).

Table 2. Factors for calculation of recharge values [20].

<table>
<thead>
<tr>
<th>Soil permeability</th>
<th>Range</th>
<th>Factor</th>
</tr>
</thead>
<tbody>
<tr>
<td>Very slow</td>
<td>&lt; 500</td>
<td>1</td>
</tr>
<tr>
<td>Slow</td>
<td>500-700</td>
<td>2</td>
</tr>
<tr>
<td>Moderate</td>
<td>700-850</td>
<td>3</td>
</tr>
<tr>
<td>High</td>
<td>&gt; 850</td>
<td>4</td>
</tr>
</tbody>
</table>

3.1.3 Impact of the vadose zone (I)

Various studies conducted on different lithologies showed that rocks containing carbonate minerals (dolomites, limestone, calcite, etc.) have a high acid neutralising potential as compared to lithologies without these minerals e.g. silicate rocks ([16]; [26]). The rate of acid neutralising potential of lithologies with silicate minerals is dependent on the silicate content, that is the less the silicate minerals the higher the potential [8]. Based on these findings lithologies within the study area were grouped into carbonates and silicates. The silicate bearing lithologies were further classified into their respective silicate content. The vadose zone impact layer (Figure 3c) shows classification of rocks in terms of their acid neutralising potential. Areas with lithologies having low acid neutralising potential are more vulnerable to AMD as less or no reactivity can mean that the AMD will pass unfiltered through the vadose zone and reach the aquifer.

3.1.4 Soils layer (S)

High resolution 1: 250 000 soil data from ARC-ISCW was reclassified according to clay content. Clay materials are known to create a barrier zone restricting water and pollutant migration, thus the more the clay content, the better the barrier effect hence lessening the chances of groundwater pollution. Figure 3d shows the clay content layer of the study area. The north part is marked by low clay content as compared to other areas.
3.1.5 Topographic slope layer (T)

SRTM data of 30 m resolution was used to generate the percentage slope layer (Figure 3e). Surface flatness increases the resident time for water and pollutants to react and infiltrate. Generally the biggest portion of the study area is flat to gently sloping as marked by dark colours (Figure 3e).

3.2 DRIST model

The five layers (Figure 3) were combined in GIS using equation (2) to generate the DRIST groundwater vulnerability model. The equation is given by:

\[
DRIST\ Index = D_T D_W + R_T R_r + I_T I_r + S_T S_r + T_T T_r \tag{2}
\]

Where \( r \) is the rating of the parameter and \( w \) is the importance weight of the parameter shown on Figure 3.

The results were divided into five zones (very low, low, moderate, high and very high) using the natural breaks method [14]. The output result is a map (Figure 11a) showing the DRIST groundwater vulnerability of the study area, where grey areas are less vulnerability and the dark coloured areas as the most vulnerable areas. The central and eastern areas have high to very high groundwater vulnerability whereas the western and southern areas are low to moderate.

3.3 ANN Training and validation

Twenty five groundwater samples collected in July 2013 were analysed by Waterlab testing laboratory for cations, anions and alkalinity using Inductively Coupled Plasma Mass Spectrometry (ICP-MS), Ion Chromatography (IC) and Spectrophotometry respectively. Physical parameters; pH and TDS were recorded in the field using Multi-parameter aquameters. The analysed hydrochemistry data was used for ANN training and validation. To identify the hydrogeochemistry system, scatter plots between various ions and TDS were plotted. The sulphate scatter plot had the highest correlation value (0.8204) (Figure 4). This shows that the study area is within a groundwater sulphates enrichment hydrogeochemistry system. In order to estimate roughly how much the sulphate contributes to the TDS values the correlation line equation was used. Theoretically if the sulphates are the only contributor to TDS then the gradient of the line would be one and be zero if it does not contribute anything. Therefore the gradient value of 0.7762 obtained (Figure 4) shows that sulphates are the major contributor and less contribution from other ions like carbonate, bicarbonate, chloride, sulphate, nitrate, sodium, potassium, calcium or magnesium. Considering this study area, the most likely process which produces sulphate into the groundwater is AMD reactions, whereby sulphide minerals found in waste rock, tailings dumps and spoil humps in abandoned or current coal mining operations on surface or in the vadose zone react with water in the presence of oxygen to form an acid solution rich in sulphate [5]. Based on these results sulphate and TDS concentrations in groundwater were selected for training of the ANN to differential areas in terms of groundwater vulnerability to AMD within the Witbank Coalfield and pH was used to validate the model.

To build an ANN system, the number of nodes in the hidden layer is very important as it controls the efficiency of the training process. According to [21] the optimum number of hidden nodes in the hidden layer is determined by multiplying the number of inputs by two and adding
one. Therefore for this study 11 nodes for the hidden layer were used. First the 25 groundwater samples were partitioned into 75% for training and 25% for validation the ANN system. The five layers of the DRIST were fed into the input layer and training program run. After training, the learnt ANN was used to predict values of the unused validation data. The difference between the predicted and actual values of the validation samples were used to establish the effectiveness of the training process and to establish the optimum number of iterations (Epochs) where the validation error is at its minimum and minimises ANN over-fitting.

### 3.2.1 Sulphate

The ANN was trained using sulphate concentration values in groundwater where areas with high sulphate values were used as polluted sites and low sulphate values as none polluted training sites. In order to get the optimum number of epochs, prediction of the partitioned validation sites was done. The results show the lowest sulphate validation Mean Square Error (MSE) of 0.075 was archived within the first 200 epochs after which the error it increases (Figure 5). Therefore the ANN was run for 200 epochs and the training errors monitored by recording the Sum of Square Errors (SSE) for each epoch. Figure 6 shows the performance of the training process with the lowest SSE value of 0.07 achieved after 200 epochs. The training results were used to predict the groundwater vulnerability values (Figure 9a) for each of the 30 m x 30 m points within the study area. The groundwater vulnerability model values was reclassified into five classes using natural break method as shown on Figure 9a. The model shows the groundwater vulnerability model generated using sulphate training where grey coloured areas are less vulnerability and the dark coloured areas as the most vulnerable areas.

### 3.2.2 Total Dissolved Solids (TDS)

The ANN was also trained separately using TDS concentration in groundwater samples. Samples with very high TDS values were used as polluted sites and very low values as none polluted training sites. The optimum number of epochs was identified by prediction the validation sites and checking the validation prediction MSE values.

![Figure 4. Scatter plot between sulphate and TDS.](image)

![Figure 5. Establishing the optimum number of epochs using sulphate validation data.](image)

![Figure 6. Performance graph of training with sulphate data.](image)

![Figure 7. Optimum number of epochs for TDS validation.](image)

![Figure 8. Performance graph of training with TDS data.](image)
The TDS values were reclassified into five classes of groundwater vulnerability zones using the natural break method (Figure 9b). The TDS model shows high values in the vicinity of the Olifants River and its tributaries. Areas near Kromdraai, Witbank and Bethal are also relatively high. The area surrounding Delmas is marked by lowest values because of the presence of high acid-neutralising rocks, high soil clay content and deeper water table.

3.3 ANN groundwater vulnerability model

The sulphate and TDS models were combined to form one model using the fuzzy (AND) operator in ArcGIS (Figure 10). The fuzzy (AND) operator was used because an area with or without AMD is deduced from both the sulphate and TDS concentrations. That is, an area with high sulphate and high TDS model values would be highly vulnerable as compared to low values for both models.

Figure 10. Fuzzy system to produce the combined model.

Figure 11b shows the combined model from the fuzzy operation illustrating the relative degree of groundwater vulnerability within the Witbank Coalfield. The model shows high values in the central to the northern parts of the coalfield.

3.4 Model Validation

The pH parameter which is a good indicator for AMD was used for model validation and for comparing the performance of the ANN and DRIST groundwater vulnerability models. Correlation scatter plots between the DRIST and ANN models with measured pH values were produced and displayed on Figure 12 (a) and (b) respectively. The pH correlation value of 0.514 was obtained for the DRIST model as compared to the 0.7182 values for the ANN model.

4. Discussion

The low pH correlation value obtained for DRIST model could be because the DRIST method is an intrinsic method which generalises the groundwater vulnerability based on the surface and subsurface conditions without considering any specific pollutant. On the other hand, the ANN model also based on the same parameters as the DRIST method but training using AMD indicators improves the pH correlation dramatically because it’s more specific to AMD pollution with which pH is a sensitive indicator. This significant improvement means that the ANN model is better in terms of highlighting the vulnerability status of the groundwater system.

The ANN model shows high groundwater vulnerability for the area surrounding the Olifants River and its tributaries. This area is marked by shallow water table, less soil clay content and overlain by rocks with poor neutralising potential showing lack of the natural attenuation of AMD which makes the groundwater highly vulnerable. The area is also marked by a high density of coal mines (Figure 1). Several authors ([23]; [3]) did studies on the Olifants River catchment and reported that high content of sulphate concentrations and low pH in both groundwater and surface water downstream of current and abandoned mining.
The pollution is reported to be affecting farmers, people living in the catchment as well as tourism and wildlife of the Kruger National Park [23]. The conditions which makes the groundwater highly vulnerable to AMD as explained in present study could be one of the reasons for results recorded by [23; 3].

In terms of possible physical chemical reasoning, the vulnerability results obtained is a combination of presence or lack of physical trapping mechanisms (i.e. Distance to water level determines the residence time for which the subsurface can attenuate AMD and interstitial void size determines AMD movement) and the soils and lithology controls the migration by pollution sorption processes, chemical reactions which are depended on the clay content and AMD-rocks reactivity.

5. Conclusion

The ANN model was build uses the DRIST input parameters (Depth to water level, Recharge, Impact of the vadose zone, Soils and Topographic slope) as inputs and hydrochemistry data (Sulphate and TDS) as training samples to produce groundwater model of Witbank Coalfield.

The results of the ANN model correlates very well with physically measured pH values as compared to the traditional DRIST model. This shows that the ANN model coupled with further field verification can be used help policy and decision makers to make scientifically informed decisions on land use planning. Based on findings of this study the following can be recommendations with regards to management and protection of groundwater resources:

- Land use activities that generate sulphates like coal mining should be avoided on highly vulnerable zones or done in a strict manner that minimise pollutants from entering the subsurface.
- Rehabilitation exercise over abundant coal mining areas which are still generating AMD within or near the highly vulnerable zones.

The datasets used in this study are readily available from various governmental agencies making the approach cost-effective in evaluating coalfield scale groundwater vulnerability to AMD. The AMD groundwater vulnerability approach developed in this research need to be applied to other pollutants with similar or different hydrogeological settings in order to determine the robustness of the methodology.
References


[18] D. Myronidis, Ch.Papageorgiou, & St. Theofanous, Landslide susceptibility mapping based on landslide history and analytic hierarchy process (AHP), Natural Hazards, 81(1), 2016, 245-263.


[22] G. Piscopo, Groundwater vulnerability map explanatory notes - Castlereagh Catchment, (Parramatta NSW; Centre for Natural Resources, NSW Department of Land and Water Conservation, 2001).


