Mine Water and the Environment (2006) 25: 214–219 © IMWA Springer-Verlag 2006

Technical Article

Prediction of Cadmium Removal Using an Artificial Neural Network and a Neuro-Fuzzy Technique

T. N. Singh[1], V.K. Singh[2], S. Sinha[2]

[1]Dept of Earth Sciences, IIT, Powai, Mumbai 400067 India; [2]Inst of Technology, Banaras Hindu University, Varanasi 221005 India; corresponding author's e-mail: tnsingh@iitb.ac.in

Abstract. The prediction of adsorption of cadmium by hematite using an adapted neural fuzzy model and a back propagation artificial neural network was compared. Adsorption was found to depend on the Cd concentration, agitation rate, temperature, pH, and the particle size of the hematite. The adaptive neuro-fuzzy inference system proved to be more efficient in predicting Cd adsorption than a single layered feed forward artificial neural network.

Key words: Adsorption; artificial neural network; cadmium; fuzzy logic; hematite

Introduction

Aquatic ecosystems can be severely affected by discharges of acidic mine water and the elevated levels of heavy metals that are often associated with them. Cadmium, in particular, is of concern: it is an irritant to the respiratory track, it can cause kidney damage, renal disorder, high blood pressure, bone degradation, and destruction of red blood cells (prolonged exposure can cause anemia) (Darsh 1993; Sitting 1981). Thus, if cadmium is present, mine water should be treated to reduce the concentration below the threshold limit value (TLV). Cadmium can be removed from water by processes such as solvent extraction, reverse osmosis, and ion exchange (Meena and Rajgopal 2003). Alkaline precipitation and ion exchange are conventionally used to reduce the concentration to a TLV of .005 mg/L (Bhattacharya et al. 1984; Patterson et al. 1977; World Health Organization 1984). Cadmium adsorption is partially diffusion controlled and partially due to an electrostatic effect, along with specific adsorption involving Cd^{++} and $CdOH^{+}$ (Singh et al. 1998). Ferric iron oxides, due to their abundance and high sorption ability, are very important natural adsorbents. Thus, an attempt has been made to develop a method for Cd removal by adsorption using hematite ore, which would be economical and less time consuming than more conventional methods. The mass transfer coefficient, diffusion coefficient, and thermodynamic parameters have been determined to elucidate the mechanism of adsorption of Cd by hematite. In the present work, a neural network model was used to predict Cd adsorption over time, based on the cadmium concentration, agitation rate, pH, and temperature.

Materials and Method

The absorption experiments were carried out using a completely mixed batch reactor (APHA et al. 1980; Bhattacharya et al. 1984; Singh et al. 1998). Similar datasets were also taken for training the network from the published literature (Bhattacharya et al. 1984; Singh et al. 1998). A sample of hematite ore from Goa, India was crushed in a ball mill to make its size 200 mesh. A solution was prepared with known amount of cadmium, maintaining a pH of 9.2 with a $NaOH/CH_3COOH$ buffer solution. The contents, at the desired concentration and pH, were agitated at different temperatures (rotary thermostat) with various sizes of hematite particles at different agitation rates. At the end of predetermined time intervals, the contents were centrifuged at 10^4 rpm and the supernatant liquors were analyzed using atomic absorption spectrophotometry (APHA et al. 1980). The laboratory tests are time consuming and expensive, and so a neural network model is proposed.

Artificial Neural Networks

An artificial neural network (ANN) is a closely interconnected network of processing units, which offers a surprisingly rich structure, exhibiting some features of a biological neural network (Yegnarayan 1997). The large number of interconnection leads to efficient predicting capabilities. The processing units constituting an ANN are called neurons. These connections are used to optimize the relative weights when training a network. A three-layered ANN is shown in Figure 1. It consists of an input layer, an output layer, and an intermediate hidden layer, with their corresponding neurons. Each neuron of an individual layer is connected with each neuron of the next layer, giving rise to a large number of connections. This enables ANNs to learn complicated patterns. Each connection has a weight associated with it. The hidden layer learns to recode (or to provide a representation for) the inputs. More than one hidden layer can be used. The architecture is more powerful than a single layer network; it can be shown that any

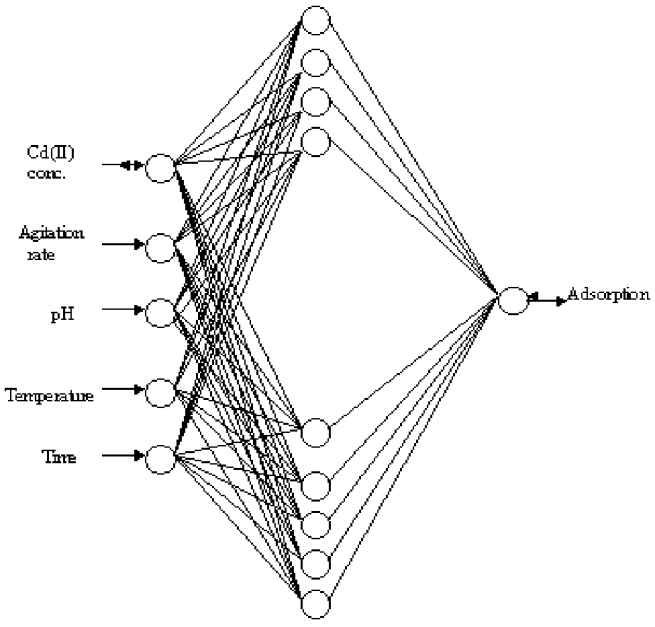

Figure 1. A 3-layered feed forward back propagation artificial neural network

mapping can be learned given two hidden layers (of units).

Training of the Network

Training of the network is basically a process of arriving at an optimum weighting. The descent down the error surface is made using the following rule:

$$\Delta w_\eta = \eta(\frac{\partial E}{\partial w_\eta}) \quad (1)$$

where η is the learning rate parameter and w_η is the weight of the connection between the i^{th} neuron of the input layer and the j^{th} neuron of the hidden layer. The update of weight for the $(n+1)^{th}$ pattern is given as:

$$w_\eta\ (n+1) = w_\eta\ (a) + \Delta\ w_\eta\ (n) \quad (2)$$

Similar logic applies to the connection between the hidden and output layers. The error E is the mean squared error and is determined by the following relation:

$$E = \Sigma\ [O_k\ (n) - O_k'\ (n)\ ]^2 \quad (3)$$

where, O_k (n) is the output determined by the network for the n^{th} pattern and O_k' (n) is the corresponding output given in the training data set.

The weight change rule is a development of the perception learning rule. Weights are changed by an amount proportional to the error at that unit times the output of the unit feeding into the weight. The output unit error is used to alter weights on the output units. Then the error at the hidden nodes is calculated (by back – propagating the error at the output units through the weights), and the weights on the hidden nodes altered using these values. For each data pair to be learned, a forward and backward pass is performed. This is repeated until the error is low enough (or the effort is abandoned).

The input and the hidden layers consist of linear processing units as neurons, whereas the output layer consists of non-linear processing units as the neurons. The non-linear function used is the logarithmic sigmoid function and is defined as:

$$f(net) = \frac{1}{(1+e^{-net})} \quad (4)$$

where (net) is the weighted sum of the inputs for a processing unit. Thus, the outputs are determined for each epoch, the mean square error calculated and the weights updated till a user specified error goal or epoch goal is reached successfully.

A three-layered, feed forward-back propagation ANN model was used. The input, output, and hidden layers had five, one, and twenty neurons respectively. The ANN was trained for 3000 epochs. The training database had 15 datasets. The testing and training database were chosen by random classification to cover the complete range of input data.

Hybrid Neural Fuzzy Model

Conventional mathematical analysis is difficult when there are uncertain and partially defined systems with a degree of vagueness. The use of intelligent systems to evaluate such complex system is ideal. An adaptive network consists of a structure of nodes and the directional links through which these nodes are connected. The outputs of the nodes depend on the parameters associated with these nodes. The change in these parameters is brought about by a learning rule to minimize the error. Optimized membership functions give more specific rules and hence increase the predicting power of the system.

Fuzzy logic is an extension of multi-valued logic. Fuzzy propositions are statements that posses fuzzy variables. The concept of a fuzzy set is the basis of fuzzy logic. A fuzzy set is a set without a crisp, clearly defined boundary. It can contain elements with only a partial degree of membership. A fuzzy set is defined by assigning each individual variable a value between 0 and 1, called its membership value. Fuzzy sets describe vague concepts and admit the possibilities of partial membership in it. The degree an object belongs to a fuzzy set is denoted by a membership value between 0, denoting absolute uncertainty and 1, denoting complete certainty. A

membership function (MF) is a curve that defines how each point in the input space is mapped to membership value (or degree of membership) (Zadeh 1965, 1979). A membership function associated with a given fuzzy set maps an input value to its approximate membership value. The mathematical form of a fuzzy set will be

$$S = \{a_s(X_1), a_s(X_2), a_s(X_3).....a_s(X_N)\} \qquad (5)$$

where $a_s(X_i)$ is the membership grade of the variable X_i, X_i belonging to the set X consisting of N number of parameters, and is defined as

$$a_s(X_i) a[0,1] i = 1,2,3,.....N \qquad (6)$$

Fuzzy interface is the process of formulating the mapping from a given input to an output using fuzzy logic. The mapping then provides a basis from which decisions can be made, or patterns discerned. The process of fuzzy interface involves membership functions, fuzzy logic operators, and if-then rules. The basic structure of the type of fuzzy interface system is a model that maps input characteristics to input membership functions, input membership functions to rules, rules to a set of output characteristics, output characteristics to output membership functions, and the output membership functions to a single-valued output or a decision associated with the output. Of the many types that the membership function might belong to, such as D-shaped curve membership function, sigmoid curve membership function and trapezoidal membership function, the output membership function used was Gaussian curve membership function. The symmetric Gaussian function depends on two parameters o and c, as given by:

$$f(x,\sigma,c) = e^{\frac{-(x-c)^2}{2\sigma^2}}$$

For utilizing the fulsome predicting power of fuzzy systems to obtain stability criteria, as in this case, effective methods are needed for the optimization of membership functions to minimize the output error measure or to make best use of the fuzzy interface system (FIS). Use of adaptive neural network techniques with FIS serves as a basis for the optimization of the appropriate membership functions to generate the stipulated input output pairs. In most cases, the fuzzy interface systems used have rules structures that are pre-stated by the programmer's understanding of the characteristics of the variables in use. The membership functions depend on the parameters, which can be however tuned to give best results by the use of an adaptive network. With the use of such neuro-adaptive techniques, the membership function parameters are optimized using a hybrid of a back propagation algorithm and a least squares methods (Demuth and Beale 1994).

Adaptive Neuro-Fuzzy Inference System

An adaptive network is a multiplayer feed forward network in which each node performs a particular function (also known as node function) on the node inputs as well as a set of parameters pertaining to this node (Jang 1993). The form for the network functions may vary from node to node, and the choice of each node function is optimized by the input-output pairs. For a given adaptive network with L layers and k^{th} layer with #(k) nodes, the node in the i^{th} position of the k^{th} layer by (k, I) and its node function by equation 7

$$O_i^k = O_i^k(O_i^{k-1},.....O_{\#k-1}^{k-1}, a, b, c.....) \qquad (7)$$

where a, b, c, etc. are the parameters associated with this node. Assuming the given training data set has P entries, the error measure for the p^{th} ($1 \le p \le P$) entry of the training data entry is the sum of squared errors:

$$E_p = \sum_{m=1}^{\#(l)} (T_{m,p} - O_{m,p}^L)^2 \qquad (8)$$

where $T_{m,p}$ is the m^{th} component of the p^{th} target output vector, and $O_{m,p}$ is the m^{th} component of the actual output vector produced by the presentation of the p^{th} input vector.

Hence, the overall error measure is $E = \sum_{p=1}^{p} E_p$

For the development of a learning procedure that implements gradient descents in E over the parameter space, the error rate becomes

$$\frac{\partial E_p}{\partial O_{l,p}^l} = -2(T_{l,p} - O_{l,p}^L) \qquad (9)$$

For the internal node at (k,i), the error rate is derived using the chain rule:

$$\frac{\partial E_p}{\partial O_{i,p}^k} = \sum_{m=1}^{\#(k+1)} \frac{\partial E_p}{\partial O_{m,p}^{k+1}} \frac{\partial O_{m,p}^{k+1}}{\partial O_{i,p}^k} \qquad (10)$$

where, $1 \le k \le L-1$. The error rate of an internal node can be expressed as a linear combination of the error rates of the nodes in the next layer. Therefore, for all $1 \le k \le l$ and $1 \le k \le \#(k)$, we can find $\frac{\partial E_p}{\partial O_{i,p}^k}$ by (3) and (4). If α is the parameter of the given adaptive network, then:

$$\frac{\partial E_p}{\partial \alpha} = \sum_{O^* \in S} \frac{\partial E_p}{\partial O^*} \frac{\partial O^*}{\partial \alpha} \quad (11)$$

where S is the set of nodes whose output depends on α. Then the derivative of the overall measure with respect to α is

$$\frac{\partial E}{\partial \alpha} = \sum_{p=1}^{p} \frac{\partial E_p}{\partial \alpha} \quad (12)$$

Accordingly, the updated formula for the generic parameter α is

$$\Delta\alpha = -\eta \frac{\partial E}{\partial \alpha} \quad (13)$$

In which η is a learning rate, which can be further, expressed as

$$\eta = \frac{k}{\sqrt{\sum \alpha (\frac{\partial E}{\partial \alpha})^2}} \quad (14)$$

where k is the step size, the length of each gradient transition in the parameter space. The method is optimized by back propagation of the algorithm with bar-delta role (Demuth and Beale 1994).

Results and Discussion

Our experimental work indicates that Cd adsorption depended on: Cd concentration, temperature, pH, agitation rate, and the particle size of the hematite. The effect of each parameter was observed with respect to time, while keeping other parameters constant (Table 1). This large dataset and data from the literature was used to develop an intelligent network to determine the non-linear behavior among these variables.

Observed and predicted values are compared in Figures 2 and 3. Figure 2 indicates a correlation between experimental values and predicted values of Cd adsorption capabilities using ANN. It exhibits very good correlation, except at a few points. The maximum error observed was 20%. On a 15 test dataset, 9 had an error percentage less than 10%, and only two points show higher error percentages, i.e. more than 18%. The variation in the surface charge is caused by adsorption and desorption of protons. The adsorption of cation and anions on the metal hydroxides and oxides surface is strongly influenced by their variable surface charge. All metal hydroxides and oxides have a surface charge that varies with pH.

Figure 3 illustrates very close correlation between observed and predicted values using ANFIS. The majority (all but two) of the observation points have an error level less than 5%. This indicates that ANFIS is better at predicting the data set than ANN. The surface diagram also shows similar behaviour in Figures 6-8. Comparative analysis between ANN and ANFIS is shown in Figure 5.

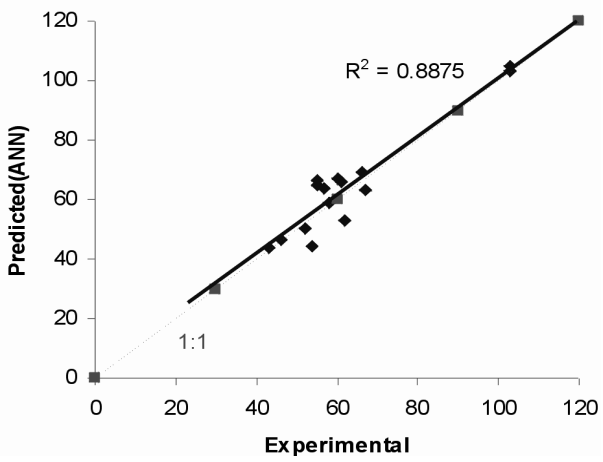

Figure 2. A graph of experimental values and values predicted by ANN

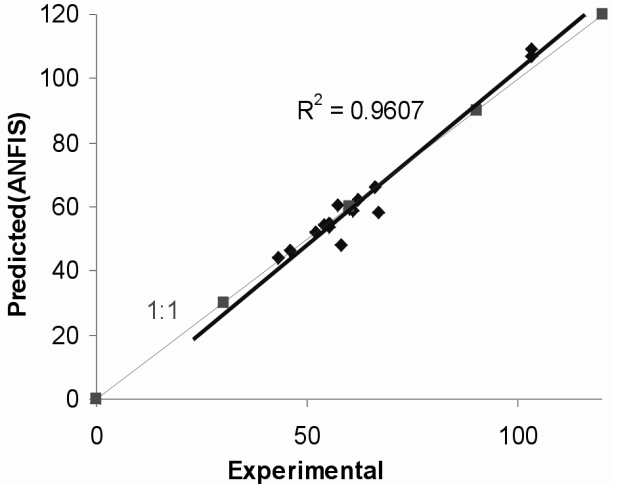

Figure 3. A graph between experimental values and values predicted by ANN

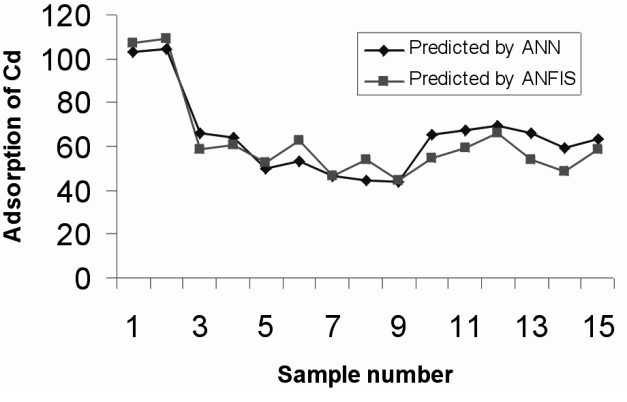

Figure 4. Comparison of results predicted by ANFIS and ANN

Table 1. Testing values of different input and output parameters and predicted adsorption values

S.No.	Cd conc µmol L^{-1}	Agitation rpm	pH	Temp. ^{o}C	Time Min	Observed Absorption µmol L^{-1}	ANN prediction µmol L^{-1}	ANFIS prediction µmol L^{-1}	% error (ANN) µmol L^{-1}	% error (ANFIS) µmol L^{-1}
1	71.17	125	9.2	40.5	222	103	103.15	106.9531	-0.1451	3.6961
2	71.17	125	9.2	40.5	127	103	104.797	109.427	-1.7442	5.874
3	88.96	125	9.2	20.5	33	61	66.1753	58.638	-8.484	-4.0282
4	71.17	125	9.2	20.5	39	57	63.9655	60.6183	-12.2202	5.969
5	62.27	125	9.2	20.5	39	52	50.0257	52.2948	3.7968	0.5637
6	62.27	125	9.2	20.5	53	62	53.0571	62.3739	14.4241	0.5994
7	53.38	125	9.2	20.5	40	46	46.7472	46.2464	-1.6244	0.5327
8	53.38	125	9.2	20.5	53	54	44.251	54.2082	18.0538	0.3841
9	44.48	125	9.2	20.5	50	43	43.943	44.2244	-2.1931	2.7686
10	71.17	50	9.2	20.5	40	55	65.1223	54.6764	-18.4042	-0.5918
11	71.17	75	9.2	20.5	39	60	67.0971	59.4698	-11.8286	-0.8916
12	71.17	75	9.2	20.5	45	66	69.3485	65.9554	-5.0735	-0.0676
13	71.17	100	9.2	20.5	30	55	66.2673	53.6896	-20.486	-2.4406
14	71.17	125	9.2	20.5	29	58	59.0288	48.3139	-1.7738	-20.0482
15	71.17	125	9.2	20.5	37	67	63.0959	58.3298	5.827	-14.8641

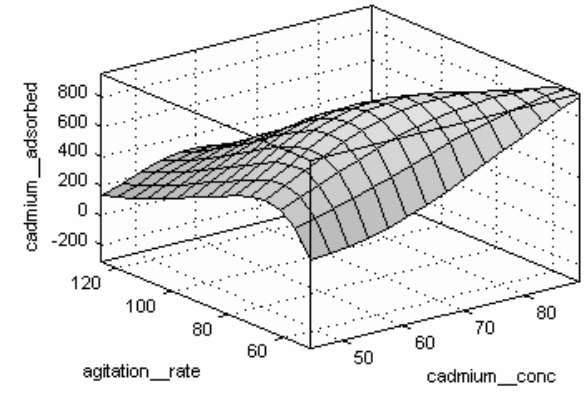

Figure 5. Surface comparison between input parameter agitation rate, cadmium concentration, and output value as predicted by ANFIS

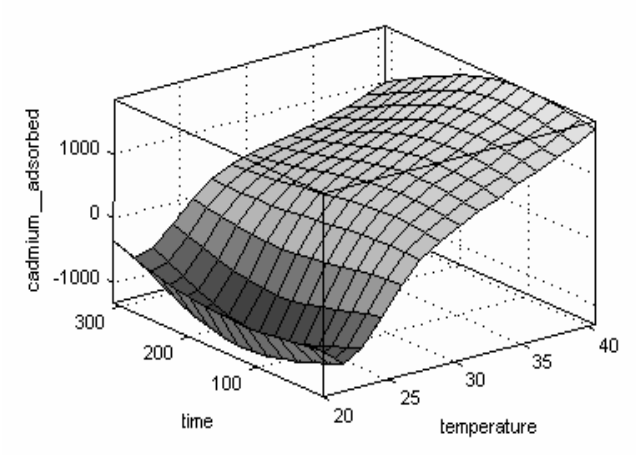

Figure 6. Surface comparison between input parameter temperature, time, and output value as predicted by ANFIS

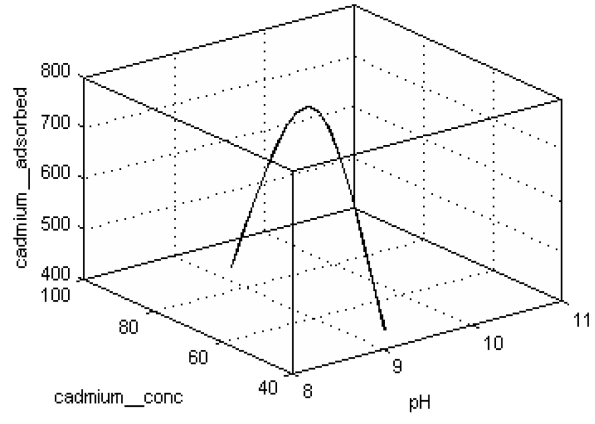

Figure 7. Surface comparison between input parameter cadmium concentration, pH and output value as predicted by ANFIS

Conclusion

The present study indicates that the hybrid neuro-fuzzy modelling method has the potential to predict the adsorption of cadmium with better accuracy than ANN, and that ANFIS captures the irregularities and uncertainties better than ANN. This predictive model can be used elsewhere, saving time and experimental effort.

References

APHA, WPCF, AWWA (1980) Standard Methods for the Examination of Water and Wastewater. 15th edit Washington, DC, 207 pp

Bhattacharya AK, Venkobacher C (1984) Removal of Cd (II) by low cost adsorbents. J Environ Eng 110: 110-122

Darsh GA (1993) Increase of cadmium body burden for this century. Sci Total Environ 67: 75-89

Demuth H, Beale M (1994) Neural Network Tool Box User's Guide. Mathworks Inc, Natick, MA

Jang JSR (1993) ANFIS: Adaptive Neural based Fuzzy Interference System. IEEE Trans on Systems, Man and Cybernetics 23: 665-685

Meena A, Rajgopal C (2003) Comparative studies on adsorption removal of chromium from contaminated water using different adsorbents. Indian J Chem Technol 10: 72-78

Patterson JW, Allen HE, Scala JJ (1977) Carbonate precipitation for heavy metal pollutants. J Water Pollut Control Fed 49: 2397

Singh DB, Rupainwar DC, Prasad G, Jayaprakas KC (1998) Studies on Cd (II) removal from water by adsorption. J Hazard Mater 60: 29-40

Sitting M (1981) Handbook of Toxic and Hazardous Chemicals. Noyes Publ, Park Ridge, NJ, USA, p 119-120, 185-186

World Health Organization (1984) Guidelines for Drinking Water Quality. Geneva, Switzerland, Vol 1 and 2

Yegnarayan B (1997) Artificial Neural Networks. Prentice-Hall of India Pvt Ltd, New Delhi, 117-135

Zadeh LA (1965) Fuzzy Sets. IEEE Information and Control 8: 338-353

Zadeh LA (1979) A Theory of Approximate Reasoning. In: Hayes JE, Michie D, Mikulich LI (Eds), Machine Intelligence 9, John Wiley & Sons, NY, p 149-194

Submitted June 18, 2004; accepted September 13, 2006

Mine Water Impressions

Chris Wolkersdorfer/Germany

Cleaning of one of the three settlement ponds for active mine water treatment of the abandoned Straßberg/Harz fluorite mine.

Mine portal of the abandoned Bad Ems/Germany "Stadt Stollen" (City Adit). This mine formerly produced iron.