Environmental Science & Technology
ARTICLE
pubs.acs.org/est

Evaluating Remedial Alternatives for an Acid Mine Drainage Stream: A Model Post Audit

Robert L. Runkel,*[,†] Briant A. Kimball,[‡] Katherine Walton-Day,[§] Philip L. Verplanck,[||] and Robert E. Broshears[⊥]

[†]U.S. Geological Survey, Mail Stop 415, Denver Federal Center, Denver, Colorado, 80225, United States
[‡]U.S. Geological Survey, 2329 West Orton Circle, West Valley City, Utah, 84119, United States
[§]U.S. Geological Survey, Mail Stop 415, Denver Federal Center, Denver, Colorado, 80225, United States
[||]U.S. Geological Survey, Mail Stop 973, Denver Federal Center, Denver, Colorado, 80225, United States
[⊥]U.S. Geological Survey (currently retired); 31 S. Indiana Place, Golden, Colorado 80401, United States

Supporting Information

ABSTRACT: A post audit for a reactive transport model used to evaluate acid mine drainage treatment systems is presented herein. The post audit is based on a paired synoptic approach in which hydrogeochemical data are collected at low (existing conditions) and elevated (following treatment) pH. Data obtained under existing, low-pH conditions are used for calibration, and the resultant model is used to predict metal concentrations observed following treatment. Predictions for Al, As, Fe, H^+, and Pb accurately reproduce the observed reduction in dissolved concentrations afforded by the treatment system, and the information provided in regard to standard attainment is also accurate (predictions correctly indicate attainment or nonattainment of water quality standards for 19 of 25 cases). Errors associated with Cd, Cu, and Zn are attributed to misspecification of sorbent mass (precipitated Fe). In addition to these specific results, the post audit provides insight in regard to calibration and sensitivity analysis that is contrary to conventional wisdom. Steps taken during the calibration process to improve simulations of As sorption were ultimately detrimental to the predictive results, for example, and the sensitivity analysis failed to bracket observed metal concentrations.

INTRODUCTION

Simulation models are often used to predict the future state of environmental systems to facilitate decision making.[1,2] Water resource managers, for example, use model results when considering the effects of changing land use or climate on water quality.[3] In a typical application, the model is calibrated such that simulations reproduce the general features of an observed data set.[4] Following calibration, model inputs are modified to reflect a change to the system for which a prediction is desired (e.g., water quality after a change in land use). The model is then rerun to provide a prediction of the future state. These predictions are by definition subject to uncertainty, as the environmental conditions characterizing the future state differ from the conditions under which the model was calibrated.[5]

Given the uncertainty in model predictions, post audits have been performed in which predictions are compared with observed data that documents what actually happened in a given system. Discrepancies between predictions and observed data may then be used to quantify model prediction uncertainty. Although post audits are rarely conducted,[5,6] they provide valuable information that can lead to changes in model structure and/or data collection activities.[2,6−10]

Assessment of prediction uncertainty is particularly important when considering the effects of hard rock mining on water quality. A recent review indicated that predictions failed to accurately characterize postmining water quality in 76% of the cases considered.[11,12] These predictions rely on various field, laboratory, and modeling techniques which quantify the processes affecting water quality. Many of these techniques are also applicable to the problem of mine site remediation. Thousands of hard rock mines are located in North America, and many mine sites are in need of remedial action.[13−16] Adits at some mines release acid mine drainage to the receiving streams, resulting in high metal concentrations and acid waters. Due to the chronic nature of the problem, the most effective remedial option often entails active treatment, wherein the adit drainage is chemically treated to reduce acidity. Because of the high cost of active treatment, there is a need for tools that help determine if a given remedial strategy

Received: July 29, 2011
Accepted: November 10, 2011
Revised: November 8, 2011
Published: November 10, 2011

ACS Publications
This article not subject to U.S. Copyright.
Published 2011 by the American Chemical Society

dx.doi.org/10.1021/es2038504 | *Environ. Sci. Technol.* 2012, 46, 340–347

will allow for the attainment of water quality standards. In response to this need, a reactive stream transport model known as OTEQ has been developed.[17] To date, OTEQ has been used to predict the effectiveness of hypothetical and actual treatment systems.[4,18] Despite the considerable amount of information provided by these applications, little is known about the uncertainty associated with model predictions. A post audit for OTEQ and the case of active treatment is therefore presented herein.

Calibration of OTEQ is conducted using a detailed synoptic data set that provides a hydrogeochemical snapshot of existing conditions.[13] For the case of active treatment, the calibrated model describes the state of the system under existing, low-pH conditions. Model predictions, in contrast, aim to predict postremediation conditions at elevated pH.[19] Because pH controls many chemical reactions, the processes controlling metal concentrations may be fundamentally different under existing and postremediation conditions. This structural change in the system leads to considerable uncertainty in model predictions.[5]

The objectives of this paper are to (1) quantify the accuracy of predictions for postremediation water quality; (2) identify sources of uncertainty in predicting postremediation water quality; and (3) identify changes in model structure and/or data collection activities that reduce prediction uncertainty. These objectives are addressed using data from Mineral Creek, an acid mine drainage stream in southwestern Colorado.

■ MATERIALS AND METHODS

Overview. Application of reactive transport models requires estimates of streamflow and solute concentration. An approach used in acid mine drainage streams is to combine the tracer-dilution method with synoptic sampling.[4,13,20−24] The tracer-dilution method provides estimates of streamflow,[25] and synoptic sampling provides a spatial description of instream and inflow chemistry. In this study, a paired synoptic approach is employed to assess prediction uncertainty. Under the paired synoptic approach, two synoptic sampling campaigns were conducted in August 2005. The initial synoptic campaign documented instream water quality under existing conditions. A second synoptic campaign documented water quality during a pH-modification experiment that elevated instream pH. This pH modification was designed to mimic the effects of an active treatment system in which a strong base is used to reduce acidity. A conservative tracer was injected during both campaigns to allow for the calculation of streamflow using the tracer-dilution method. The two sampling campaigns are denoted herein as MC1 and MC2.

Assessing prediction uncertainty is a three-step process (Figure 1). In the calibration step (step 1), MC1 data are used to develop a model that describes the hydrogeochemical processes that control concentrations under existing conditions. During calibration, model inputs are adjusted to obtain a close correspondence between simulated and observed concentrations. Following calibration, the simulation reproduces the features of the data by considering the loading attributed to source waters and the geochemical processes that act to attenuate concentrations. In the prediction step (step 2), the calibrated model is used to simulate postremediation water quality. In this simulation, chemistry at the upstream boundary of the system is modified to reflect the effects of active treatment. The resultant simulation provides estimates of postremediation water quality that reflect the reduction in acidity and metal concentrations afforded by the treatment system. In the assessment step (step 3), postremediation results are compared with MC2 data to quantify prediction uncertainty. This comparison is used to identify changes in model structure and/or data collection activities that would reduce prediction uncertainty.

Implementation of steps 1−3 follows the "blind" approach,[3,8] with the lead author acting as the modeler, performing steps 1 and 2 without knowledge of the MC2 results. Data are released from the blind (as held by the coauthors) following step 2. Assessment of prediction uncertainty (step 3) is then made by the entire research team.

Tracer Injection, Base Addition, and Synoptic Sampling. Continuous injections for both synoptic campaigns were initiated at the upstream end of the study reach. The conservative tracer was prepared by adding LiBr to streamwater; the strong base was prepared by mixing a 50% NaOH solution with water. The conservative tracer injection for the initial synoptic campaign was conducted August 21−22; tracer injections for the second synoptic campaign were conducted August 23−24. Synoptic samples were collected at 24 stream sites and 11 inflow locations (Figure 2) on August 22 and August 24. Total recoverable and dissolved cation concentrations were determined from unfiltered and filtered samples. Estimates of streamflow were determined from tracer dilution.[20,25]

Reactive Transport Modeling. OTEQ is formed by coupling a solute transport model[26] with a chemical equilibrium submodel (MINTEQ[27]). The coupled model considers advection, dispersion, transport and deposition of water-borne solid phases, acid/base reactions, complexation, precipitation/dissolution, and sorption.[17] Governing equations are in terms of components, where the total concentration is the sum of dissolved, precipitated, and sorbed species. Partitioning between dissolved, precipitated, and sorbed phases is based on submodel calculations for each stream segment.

Components include Al, Ca, Cd, Cu, CO_3, F, Fe(II), Fe(III), H_3AsO_4, Mg, Mn, Ni, Pb, SO_4, TOTH (total excess hydrogen), and Zn. Precipitation reactions are defined for microcrystalline gibbsite ($Al(OH)_3$) and ferrihydrite ($Fe(OH)_3$). Sorption of As,

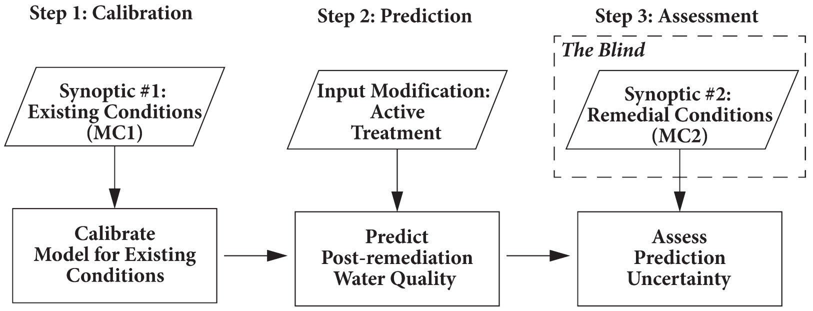

Figure 1. Assessment of model prediction uncertainty under the paired synoptic approach.

Ca, Cd, Cu, Ni, Pb, SO_4, TOTH, and Zn to precipitated $Fe(OH)_3$ is modeled using a surface complexation approach and database.[27−29] The mass of sorbent within each stream segment is based on the amount of $Fe(OH)_3$ within the water column, as determined by the submodel.

Unless noted otherwise, equilibrium constants for reactions are set using default values from the equilibrium submodel.[17,27,29] Sorption parameters (i.e., surface area, sorbent molecular weight, low affinity site density) are set using the best estimates of Dzombak and Morel.[29] High affinity site density is set equal to the upper reported value,[29] reflecting the high sorptive capacity of freshly precipitated $Fe(OH)_3$.[28]

Calibration: Existing Conditions. Estimates of streamflow, hydrologic parameters, and inflow chemistry based on the MC1 data were used to develop a calibrated model of existing conditions. The study reach was divided into 20 reaches based on changes in streamflow and water chemistry (Supporting Information (SI) Table S1) and the upstream boundary condition was set using component concentrations of the sample collected at 576 m. Component concentrations at the upstream boundary and within the inflows were set equal to total recoverable concentrations for most components. Two exceptions are TOTH and CO_3, components that were assigned concentrations based on stand-alone MINTEQ computations. In these computations, pH and alkalinity were fixed at observed values, and TOTH and CO_3 were determined from the speciated output. For samples without alkalinity, CO_3 concentrations were based on equilibrium with atmospheric CO_2.

Assignment of inflow chemistry based on sampled inflows (SI Table S1) and the use of default equilibrium constants results in a simulation that has not been calibrated to match observed stream data. This simulation reproduces the features of the observed concentration profiles but is lacking in some respects. Several modifications were therefore made to improve the calibration (see SI).

Prediction: Postremediation Water Quality. Predictive simulations were performed without prior knowledge of MC2 results. Prediction of postremediation water quality required two changes to the calibrated model. First, MC1 streamflow estimates were replaced with estimates from MC2 (streamflow decreased ~17% from August 22 to August 24).[20] Second, the upstream boundary condition was modified to consider the effects of the pH modification. MC2 values of pH and alkalinity at 576 m were therefore released from the blind and used within a series of MINTEQ runs to develop revised concentrations of TOTH, CO_3, Fe(II), and Fe(III) at the upstream boundary. Within these runs, NaOH was added to the MC1 solution composition at the upstream boundary and gibbsite and ferrihydrite were allowed to precipitate. Final values of TOTH, CO_3, Fe(II), and Fe(III) were obtained by adding enough NaOH to reproduce the observed pH and by redistributing Fe(II) and Fe(III) concentrations consistent with the calibrated model. Boundary concentrations for the remaining components were unchanged from their MC1 values. Given these changes to the calibrated model, a simulation was performed to obtain the modeler's "best estimate" of postremediation water quality.

The approach to estimating postremediation water quality is consistent with previous efforts[4,18,24,30] in which simulations were designed to "err on the conservative side" (i.e., to select inputs that are expected to underestimate removal). Remedial actions are thus expected to result in water quality improvements that are at least as large as those predicted by the model. Simulations are expected to err on the conservative side for two reasons. First, precipitation reactions are limited to the formation of gibbsite and ferrihydrite. Other precipitates that form at elevated pH are not modeled (e.g., $CdCO_3$, $ZnCO_3$), and the simulations overestimate dissolved concentrations for components affected by the formation of these precipitates. Second, the simulations do not consider gas exchange reactions in which the CO_3 concentration is in equilibrium with atmospheric CO_2. Gas exchange would result in a higher simulated pH that would in turn increase the extent of precipitation and sorption reactions.

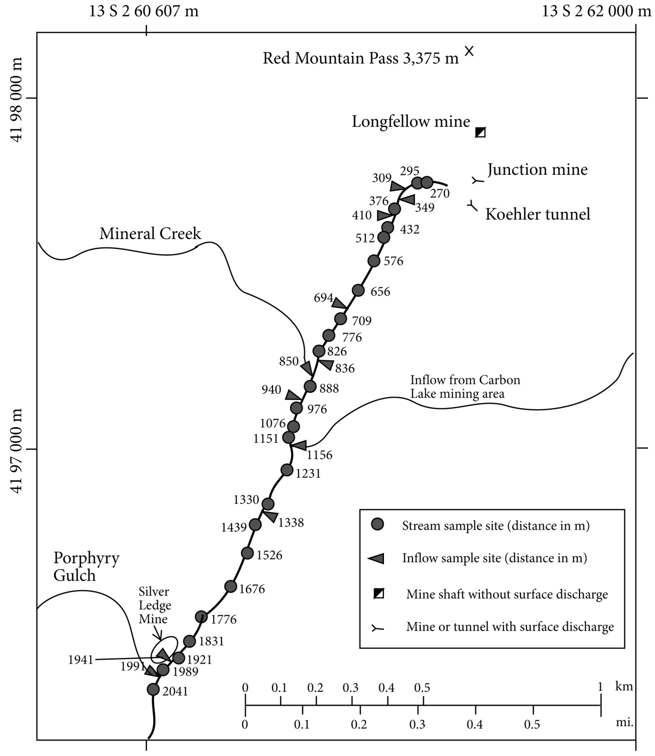

Figure 2. Mineral Creek study reach with stream and inflow sampling sites.

To further quantify prediction uncertainty, a series of simulations were performed to supplement the best estimate of postremediation water quality. These simulations employed a range of inputs in an effort to develop upper and lower bounds on postremediation concentrations.[24] These simulations were intended to form an envelope around the observed MC2 concentrations. The additional simulations were conducted within a factorial design framework[31] in which M variables are evaluated at N levels (N^M simulations). Three variables were analyzed at two levels, leading to eight simulations (SI Table S3), where

- **Variable 1: Gas Exchange.** OTEQ options for considering gas exchange include no exchange, as in the case of the best estimate simulation, and equilibrium with atmospheric CO_2.[17] These options constitute two levels within the factorial design.
- **Variable 2: Solid phases at the Upstream Boundary.** The increase in pH results in the formation of solid phases through precipitation and sorption reactions. The presence or absence of solid-phase material within the water column at the upstream boundary has potential implications for reactions further downstream; for example, the amount of precipitated Fe determines the amount of sorbent mass and thus affects dissolved concentrations of other components.

The best estimate simulation specifies boundary concentrations that are equal to the MC1 total recoverable concentrations, under the assumption that solid phases produced by the elevated pH remain in the water column (level 1). An alternative assumption is that solid-phase material settles out before reaching the upstream boundary. This possibility is investigated by setting the boundary concentrations equal to the dissolved concentrations from the MINTEQ run for the upstream boundary (level 2).

- **Variable 3: Surface Complexation Constants.** OTEQ simulations to date[4,18,24,30] have used the surface complexation constants of Dzombak and Morel.[29] Alternate values of surface complexation constants have been proposed by other investigators,[32−35] however. Simulations are therefore performed using the default values[29] (level 1) and the surface complexation constants proposed by Balistrieri et al.[34] (level 2).

Assessment. Prior to conducting the predictive simulations, a procedure for quantifying prediction uncertainty was outlined by the lead author. This a priori procedure calls for the evaluation of percent relative error at three points of compliance. These points of compliance were arbitrarily placed at sites immediately above large tributary inflows (826, 1151, and 1989 m). Percent Relative Error (PRE) is given by

$$\mathrm{PRE} = 100\frac{C_{\mathrm{MC2}} - \hat{C}_{\mathrm{MC2}}}{C_{\mathrm{MC2}}} \tag{1}$$

where C_{MC2} and $\hat{C}_{\mathrm{MC2}}$ are the observed and simulated concentrations of dissolved constituents following remediation.

Two a posteriori metrics for quantifying prediction uncertainty were formulated following release of the MC2 data set. The percent reduction (PR) is given by

$$\mathrm{PR} = 100\frac{C_{\mathrm{MC1}} - c_{\mathrm{MC2}}}{C_{\mathrm{MC1}}} \tag{2}$$

where C_{MC1} is the observed concentration under existing conditions and c_{MC2} is either the observed (C_{MC2}; when calculating observed percent reduction) or simulated ($\hat{C}_{\mathrm{MC2}}$; when calculating simulated percent reduction) dissolved concentration following remediation. A second a posteriori metric, the Water Quality Standard Assessment (WQSA), compares the observed and simulated concentrations with applicable water quality standards (chronic aquatic-life standard[36]) to determine whether or not the simulation correctly identifies attainment or nonattainment (i.e., are observed and simulated concentrations on the same side of the water quality standard?). Application of WQSA is analogous to the identification of Type I and II errors in hypothesis testing. Type I errors (false positive) arise when simulation results incorrectly indicate the attainment of water quality standards, whereas Type II errors (false negative) arise when simulation results incorrectly indicate nonattainment (Table 1). In this context, Type I errors are more severe than Type II errors, as the proposed remedial action will fail to meet the water quality standard and additional treatment may be required.

Table 1. Water Quality Standard Assessment (WQSA)

		Observed Data (MC2)	
		Attainment	Non-attainment
Best Estimate Simulation	Attainment	Correct Assessment	Type I Error: False Positive
	Non-attainment	Type II Error: False Negative	Correct Assessment

Assessment of prediction uncertainty using eqs 1 and 2 and WQSA is based on dissolved concentrations, consistent with the water quality standards; simulated concentrations are taken from the best estimate simulation. Observed concentrations that fall below analytical detection limits are set equal to one-half the detection limit within eqs 1 and 2 and the figures that follow (C_{MC2} values for Fe and Pb).

RESULTS AND DISCUSSION

Existing Conditions. Mine drainage results in acidic, metal-rich waters in the upper portion of the study reach (576−826 m; Figure 3a−d, SI Figure S1a−e). Concentrations and acidity are diluted at 850 m by the circumneutral inflow. Concentrations exhibit a gradual decrease and pH exhibits a gradual increase over the remainder of the study reach (888−2041 m; Figure 3a−d, SI Figure S1a−e). Low pH inhibits precipitation and sorption reactions, such that Al, Cd, Cu, Mn, and Zn are subject to conservative transport (total recoverable ≈ dissolved; SI Figure S1a−e). Arsenic, Fe, and Pb, in contrast, are subject to precipitation and sorption reactions that result in the formation of solid phases (total recoverable > dissolved; Figure 3b−d). Concentrations of Al, Cd, Cu, Fe, H^+, Mn, Pb, and Zn exceed chronic aquatic life standards along the entire study reach. The calibrated model reproduces the features of the observed spatial profiles, including the reactive transport of As, Fe, and Pb (Figure 3a−d, SI Figure S1a−e).

Post-Remediation Water Quality. The pH increase afforded by the base injection (>6, Figure 3e) results in the formation of solid-phase material through precipitation and sorption reactions that reduce dissolved concentrations (Figure 3f−h; SI Figure S1f−j). As a result, most constituents meet chronic aquatic life standards at all (As, Fe, Figure 3f,k) or most (Al, SI Figure S1k; Cu, SI Figure S1m; Pb, Figure 3l; pH, Figure 3i) of the stream sites. Dissolved concentrations of Cd, Mn, and Zn remain above the standard at most stream sites, however, with concentrations increasing in the downstream direction (SI Figure S1l,n,o). The observed increases in concentration are attributable to desorption or dissolution of solid-phase material within the water column in response to decreasing pH with distance (Figure 3i, SI Figure S1g,i,j). These increases would probably not be present if the treatment system included a settling pond to prevent downstream transport of solid-phase material. Implementation of an active treatment system at the head of the study reach thus appears to have the potential to result in attainment of water quality standards for all constituents, with the exception of Pb (concentrations exceed the standard at the final two stream sites due to an untreated source at 1941 m,[4,21] Figure 3l).

Assessment. Best estimate simulations reproduce the general features of the total recoverable and dissolved concentration data and observed pH (Figure 3e−h; SI Figure S1f,h). Simulations of dissolved concentration appear to closely match observed concentrations at the scale presented, whereas simulations of total recoverable concentrations generally underestimate observed concentrations (Figure 3f−h; SI Figure S1f,h). Discrepancies between simulated and observed total recoverable concentrations are attributable to the use of MC1 concentrations to specify the upstream boundary condition (total recoverable MC2 concentrations were 20−346% higher than MC1 concentrations due to source variation[20,21]). These discrepancies are not of consequence,

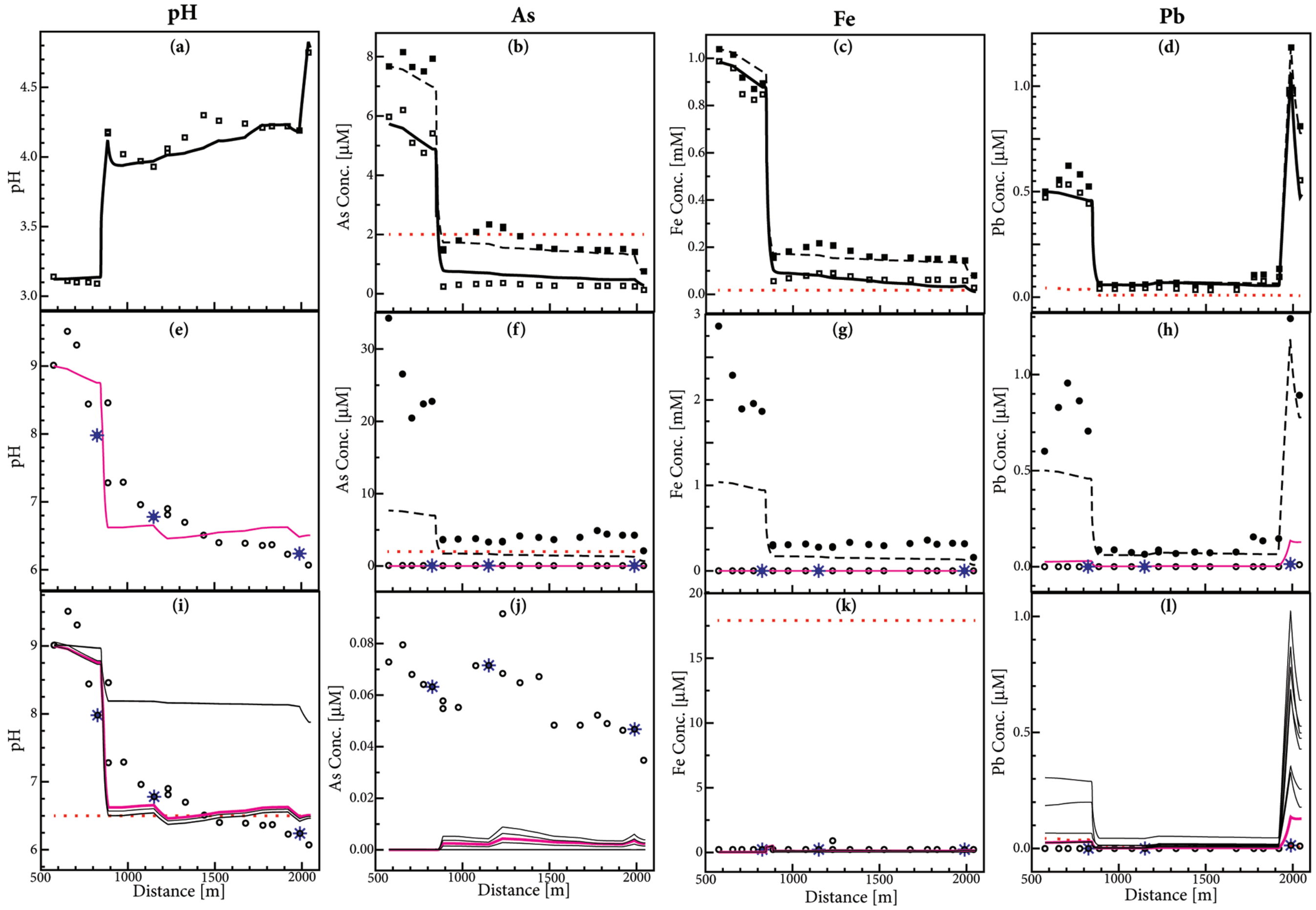

Figure 3. Data and simulations for existing conditions and remediation. Data include total recoverable and dissolved concentrations under existing conditions (filled and open squares) and remediation (filled and open circles; compliance points denoted by blue stars). Simulations include total recoverable (dashed line) and dissolved concentrations (solid line; best estimate shown in color). The chronic aquatic life standard is denoted by dotted red line. (a–d) data and simulations from model calibration, existing conditions; (e–h) data and best estimate simulation, remediation; (i–l) data and eight simulations of dissolved concentration, remediation.

however, as most water quality standards are in terms of dissolved concentration (one exception is Fe, see "Sources of Error").

Despite the agreement noted above, closer examination of the dissolved concentrations reveals substantial errors in the estimation of postremediation water quality (Figure 3j–l; SI Figure S1k,m). All eight simulations of dissolved As, for example, underestimate the observed concentrations (Figure 3j). Underestimation and overestimation of dissolved concentration is formally quantified as the Percent Relative Error (eq 1), with absolute PRE values ranging from 11 (Al) to 114 400% (Cu) (Table 2). Although the high absolute PRE values are cause for concern, this concern is tempered when PRE values are viewed in light of "erring on the conservative side". Under this concept, the preference is to avoid overestimating the reduction in dissolved concentrations (avoid underestimating MC2 concentrations), such that the effectiveness of the treatment system is not overestimated. Overestimates in the amount of reduction result in positive values of PRE and these values are generally acceptable (11–100%, Table 2). Underestimates in the amount of reduction result in negative PRE values, and these values range from −35 to −114,400%. The higher absolute values of PRE are therefore associated with underestimating the amount of reduction, and this underestimation is preferable to overestimation.

Two a posteriori metrics were formulated to supplement PRE. The percent reduction (eq 2) provides a means of comparing observed and simulated percent reduction (OPR and SPR). If OPR and SPR are in general agreement, the simulation has adequately predicted the postremediation reduction in dissolved concentration relative to existing conditions. The percent reductions that are observed and simulated for H^+, Al, and As are all comparable, for example (Table 2, SI Figure S2a–c), whereas there is a stark difference in OPR and SPR for Cd, Mn, and Zn (Table 2, SI Figure S2d,g,i). The WQSA provides a means to determine whether or not the simulations provide the correct answer in regard to attainment of water quality standards. The correct answer is obtained when the simulated concentration and observed data are both higher or lower than the applicable standard. The WQSA results indicate that the correct answer is provided in 19 of the 25 cases (green, Table 2). Of the six incorrect results, only two are Type I errors in which the simulations incorrectly predict attainment of the standard (red, Table 2), with the remaining four incorrect results being less severe Type II errors in which the simulation incorrectly indicated nonattainment (yellow, Table 2).

Although the percent reduction and WQSA metrics were formulated following the release of the MC2 data, they appear to be

Table 2. Assessment Results

	Compliance Point [m]	Percent Relative Error [%]	Percent Reduction [%][a]		Is Concentration < Chronic Standard?[b]	
			Obs.	Sim.	Obs.	Sim.
Al	826	93	97	100	🟩 Y	🟩 Y
Al	1151	65	98	99	🟩 Y	🟩 Y
Al	1989	11	98	98	🟩 Y	🟩 Y
As	826	100	99	100	🟩 Y	🟩 Y
As	1151	97	79	99	NA	NA
As	1989	92	81	99	NA	NA
Cd	826	-8,414	99	0	🟩 N	🟩 N
Cd	1151	-246	76	16	🟩 N	🟩 N
Cd	1989	29	-25	11	🟩 N	🟩 N
Cu	826	-114,400	100	7	🟨 Y	🟨 N
Cu	1151	-2,934	100	89	🟨 Y	🟨 N
Cu	1989	-148	93	82	🟩 N	🟩 N
Fe	826	88	100	100	🟩 Y	🟩 Y
Fe	1151	53	100	100	🟩 Y	🟩 Y
Fe	1989	37	100	100	🟩 Y	🟩 Y
H+	826	83	100	100	🟩 Y	🟩 Y
H+	1151	-35	100	100	🟩 Y	🟩 Y
H+	1989	43	99	99	🟩 N	🟩 N
Mn	826	-587	83	-15	🟨 Y	🟨 N
Mn	1151	21	10	28	🟥 N	🟥 Y
Mn	1989	40	-26	25	🟥 N	🟥 Y
Pb	826	-10,964	100	93	🟩 Y	🟩 Y
Pb	1151	-634	99	96	🟩 Y	🟩 Y
Pb	1989	-902	99	86	🟩 N	🟩 N
Zn	826	-76,991	100	16	🟨 Y	🟨 N
Zn	1151	-852	91	16	🟩 N	🟩 N
Zn	1989	14	-4	11	🟩 N	🟩 N

[a] "Obs." and "Sim." values are the Observed and Simulated Percent Reduction values (OPR and SPR) discussed in the main text. These OPR and SPR values are illustrated in Figure S2 of the Supporting Information.

[b] Water Quality Standard Assessment: correct predictions in green, Type I errors in red, Type II errors in yellow (Table 1); NA, constituent meets standard under existing conditions.

a better measure of prediction uncertainty than the a priori metric, PRE. Percent relative errors for Pb, for example, are high, yet there is excellent agreement between the observed and simulated percent reductions and the simulations correctly predict attainment or nonattainment of standards (Table 2, Figure 3l, SI Figure S2h). Using the two a posteriori metrics as a guide, one can conclude that predictions for Al, As, Fe, H$^+$, and Pb are superior to the predictions for Cd, Cu, Mn, and Zn (Table 2).

Sources of Error. Potential sources of error in predictive simulations noted here and elsewhere[4,24,30] include (1) identification of the appropriate solid phases and associated equilibrium constants for precipitation reactions, (2) specification of gas exchange reactions affecting CO_3, (3) specification of solid-phase Fe at the upstream boundary, and (4) specification of surface complexation constants for sorption reactions. For the application considered herein, sources 1 and 2 do not result in appreciable error. In regard to source 1, postremediation pH ranges from 6.1 to 9.5 (Figure 3e), causing Al and Fe to become insoluble. Simulations of dissolved Al and Fe are relatively insensitive to the specified equilibrium constants in this pH range and estimates of postremediation concentrations are sufficiently accurate (Figure 3g, SI Figure S1f, Table 2). In regard to source 2, the best estimate simulation does not consider gas exchange and there is an adequate reproduction of pH (Figure 3e, Table 2). Sources 3 and 4, in contrast, play a large role in causing the errors associated with the constituents subject to sorption. As noted in the Assessment subsection, simulations of total recoverable concentrations underestimate the amount of solid-phase material due to misspecification of the upstream boundary condition. Errors in simulating total recoverable Fe are of great consequence as the amount of solid-phase Fe determines the sorbent concentration. Underestimation of solid-phase Fe (Figure 3g) therefore results in the underestimation of Cd, Cu, Pb, and Zn sorption and overestimation of the corresponding dissolved concentrations (Figure 3l; SI Figure S1l,m,o). In contrast to the other constituents subject to sorption, dissolved As concentrations are underestimated (Figure 3j); overestimation of As sorption is not attributable to misspecification of solid-phase Fe, and the error is attributable to source 4.

Simulations of postremediation water quality are remarkably good in light of the limited information used to initiate the predictions (MC2 values of pH and alkalinity at the upstream boundary). Predictions for Al, As, Fe, H$^+$, and Pb accurately reproduce the observed percent reduction in dissolved concentrations afforded by the treatment system and the information needed by decision makers in regard to standard attainment is also accurate (Assessment subsection and Table 2). In addition, errors associated with most of the remaining constituents (Cd, Cu, Zn) may be attributed to the misspecification of the upstream boundary condition, a situation that could be remedied by obtaining additional information on treatment system design. Errors in the prediction of postremediation Mn are also easily explained, as Mn transport was not rigorously considered during calibration. The simulation shown in SI Figure S1d was deemed adequate and no attempt was made to increase the simulated concentration by increasing the amount of Mn entering via inflow waters (similar changes were made for other constituents; SI Table S2). If such an action were taken, OPR and SPR values would be in better agreement and the Type I errors may have been eliminated (as a result, all of the WQSA errors would be the less serious Type II). In summary, the errors in predicting postremediation water quality can be attributed to misspecification of model inputs and parameters, as opposed to inadequacies in model structure. Changes to model structure, such as more rigorous consideration of gas exchange[30] may still be warranted, however, as other applications may yield different results.

The encouraging results noted above are not surprising given the fact that simulated reactions are based on theoretically rigorous laws of mass action and an existing database of equilibrium constants.[17,27,29] The large difference in pre- and postremediation pH (3.1−4.8 versus 6.1−9.5; Figure 3a,e), however, represents a structural change to the modeled system[5] that complicates the calibration process. For the case of As sorption, calibration led to a change in the surface complexation constant (see SI) that exacerbated the prediction error (the change improved the calibration by increasing the amount of sorption, but the postremediation simulation requires less sorption, not more). The structural change altered the sorbent/sorbate ratio through the precipitation of Fe, such that the calibrated surface complexation constant was invalid at elevated pH.

Modeling Strategies. Model application is a partially subjective exercise in which ad hoc strategies are used to address study goals. Post audits provide valuable information that can be used to assess the effectiveness of these strategies, with the benefit of hindsight. The post audit provides insight into three issues that are specific to modeling reactive solute transport and modeling in general. First, one strategy employed herein is to "err on the conservative side", such that the effectiveness of the proposed treatment system is not overestimated. This strategy was largely effective as the majority of errors associated with WQSA were Type II, where model results suggest the need for additional treatment capacity. This strategy thus leans toward the attainment of water quality standards, as the simulations would lead decision makers to add unneeded treatment capacity (as opposed to Type I errors, where the proposed treatment would be deemed adequate, when in fact it is not). Decision makers should therefore balance the potential nonattainment of standards with the expense of an overdesigned treatment system.

Second, conventional wisdom dictates that sensitivity analyses provide a means to develop a range of predictions that bracket the actual outcome.[24] The factorial design does not meet this goal, as predicted concentrations from the eight simulations sometimes lie entirely above (e.g., Cu, Pb) or below (As) the data (Figure 3j,l; SI Figure S1m). The additional simulations used to perform the sensitivity analysis actually provide misleading information, as the best estimate simulation often provides the closest reproduction of the observed data (e.g., Cu and Pb, Figure 3l, SI Figure S1m). The factorial design may have been more successful if additional levels had been considered. Consideration of higher solid-phase concentrations, for example, would have lowered the concentrations of constituents subject to sorption.

Third, although calibration is an essential part of the modeling process, calibration may or may not lead to more accurate predictions. As with most predictions, structural changes in the modeled system are a cause for concern;[5] pH differences between existing conditions and remediation result in a situation in which many processes within the model are uncalibrated. Precipitation of Al, for example, was not considered during the calibration process as low pH resulted in conservative transport. This particular uncalibrated reaction was not of consequence, however, given the satisfactory results obtained for Al (Table 2). Simulation of As sorption, in contrast, was subject to calibration, and the changes made during calibration were detrimental to predictive results. A final example is the case of Mn, a constituent that was ignored during the calibration process; inclusion of this constituent in the calibration process would have improved the predictive results, possibly eliminating the only Type I errors.

ASSOCIATED CONTENT

Ⓢ Supporting Information. Additional information as noted in the text. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Phone: 303/541-3013. Fax: 303/541-3084. E-mail: runkel@usgs.gov.

ACKNOWLEDGMENT

We acknowledge the support of the USGS Toxic Substances Hydrology Program, San Juan County, San Juan RC&D, and U.S. EPA. Numerous individuals contributed to this study (see Supporting Information); helpful review comments were provided by Kirk Nordstrom and three anonymous reviewers. Opening photograph credited to P.L.V.

REFERENCES

(1) Lung, W. S. *Water Quality Modeling for Wasteload Allocations and TMDLs*; John Wiley & Sons: New York, 2001.

(2) Tiedeman, C. R.; Hill, M. C. Model calibration and issues related to validation, sensitivity analysis, post-audit, uncertainty evaluation and assessment of prediction data needs. *In Groundwater: Resource Evaluation, Augmentation, Contamination, Restoration, Modeling and Management*; Thangarajan, M., Ed.; Springer: Netherlands, 2007; pp 237−282.

(3) Ewen, J.; Parkin, G. Validation of catchment models for predicting land-use and climate change impacts. 1. Method. *J. Hydrol.* **1996**, *175*, 583–594.

(4) Runkel, R. L.; Kimball, B. A. Evaluating remedial alternatives for an acid mine drainage stream: Application of a reactive transport model. *Environ. Sci. Technol.* **2002**, *36*, 1093–1101.

(5) van Straten, G. Models for water quality management: The problem of structural change. *Water Sci. Technol.* **1998**, *37*, 103–111.

(6) Anderson, P. F.; Lu, S. A post audit of a model-designed ground water extraction system. *Ground Water* **2003**, *41*, 212–218.

(7) Lesht, B. M.; Fontaine, T. D.; Dolan, D. M. Great Lakes total phosphorus model: Post audit and regionalized sensitivity analysis. *J. Great Lakes Res.* **1991**, *17*, 3–17.

(8) Thorne, M. C.; Degnan, P.; Ewen, J.; Parkin, G. Validation of a physically based catchment model for application in post-closure radiological safety assessments of deep geological repositories for solid radioactive wastes. *J. Radiol. Prot.* **2000**, *20*, 403–421.

(9) Konikow, L. F. Predictive accuracy of a ground-water model—Lessons from a Postaudit. *Ground Water* **1986**, *24*, 173–184.

(10) Rutherford, J. C. Deoxygenation in a mobile-bed river—II. Model calibration and post-audit. *Water Res.* **1991**, *25*, 1499–1508.

(11) Kuipers, J. R.; Maest, A. S.; MacHardy, K. A.; and Lawson, G. *Comparison of Predicted and Actual Water Quality at Hardrock Mines: The Reliability of Predictions in Environmental Impact Statements*; Kuipers & Associates and Buka Environmental: Butte MT and Boulder CO, 2006.

(12) Pelley, J. Unreliable water-quality forecasts at mines. *Environ. Sci. Technol.* **2007**, *41*, 1511–1512.

(13) Kimball, B. A.; Walton-Day, K.; Runkel, R. L. Quantification of metal loading by tracer injection and synoptic sampling, 1996−2000. In *Integrated Investigations of Environmental Effects of Historical Mining in the Animas River Watershed, San Juan County, Colorado*, USGS Professional Paper 1651; Church, S. E., von Guerard, P., Finger, S. E., Eds.; U.S. Geological Survey, **2007**, pp. 417−495.

(14) Benner, S. G.; Blowes, D. W.; Gould, W. D.; Herbert, R. B.; Ptacek, C. J. Geochemistry of a permeable reactive barrier for metals and acid mine drainage. *Environ. Sci. Technol.* **1999**, *33*, 2793–2799.

(15) Shokes, T. E.; Moller, G. Removal of dissolved heavy metals from acid rock drainage using iron metal. *Environ. Sci. Technol.* **1999**, *33*, 282–287.

(16) Paulson, A. J.; Balistrieri, L. Modeling removal of Cd, Cu, Pb, and Zn in acidic groundwater during neutralization by ambient surface waters and groundwaters. *Environ. Sci. Technol.* **1999**, *33*, 3850–3856.

(17) Runkel, R. L. *One-Dimensional Transport with Equilibrium Chemistry (OTEQ): A Reactive Transport Model for Streams and Rivers; USGS Techniques and Methods Book 6, Chapter B6*, 2010. http://water.usgs.gov/software/OTEQ/ (accessed July 15, 2011).

(18) Ball, J. W.; Runkel, R. L.; Nordstrom, D. K. Reactive transport modeling at high-flow—Wightman Fork/Alamosa River, USA. In *Proceedings of the Tenth International Symposium on Water-Rock Interaction, Villasimius, Italy, July 10−15, 2001*; Cidu, R., Ed.; A.A. Balkema Publishers: Netherlands, 2001, pp 1181−1184.

(19) Broshears, R. E.; Runkel, R. L.; Kimball, B. A.; McKnight, D. M.; Bencala, K. E. Reactive solute transport in an acidic stream: Experimental pH increase and simulation of controls on pH, aluminum, and iron. *Environ. Sci. Technol.* **1996**, *30*, 3016–3024.

(20) Runkel, R. L.; Kimball, B. A.; Steiger, J. I.; Walton-Day, K. *Geochemical data for upper Mineral Creek, Colorado, under existing ambient conditions and during an experimental pH modification*, August 2005, USGS Data Series 442, 2009. http://pubs.usgs.gov/ds/442/pdf/DS442.pdf (accessed July 15, 2011).

(21) Runkel, R. L.; Bencala, K. E.; Kimball, B. A.; Walton-Day, K.; Verplanck, P. L. A comparison of pre- and post-remediation water quality, Mineral Creek, Colorado. *Hydrol. Process* **2009**, *23*, 3319–3333.

(22) Bencala, K. E.; McKnight, D. M. Identifying in-stream variability—Sampling iron in an acidic stream. In *Chemical Quality of Water and the Hydrologic Cycle*; Averett, R. C., McKnight, D. M., Eds.; Lewis Publishers: Ann Arbor, 1987; pp 255−269.

(23) Kimball, B. A.; Broshears, R. E.; Bencala, K. E.; McKnight, D. M. Coupling of hydrologic transport and chemical reactions in a stream affected by acid mine drainage. *Environ. Sci. Technol.* **1994**, *28*, 2065–2073.

(24) Runkel, R. L.; Kimball, B. A.; Walton-Day, K.; Verplanck, P. L. A simulation-based approach for estimating premining water quality: Red Mountain Creek, Colorado. *Appl. Geochem.* **2007**, *22*, 1899–1918.

(25) Kilpatrick, F. A.; Cobb, E. D. *Measurement of Discharge Using Tracers.* USGS Techniques of Water-Resources Investigations Reports, Book 3, Chapter A16; U.S. Geological Survey, 1985.

(26) Runkel, R. L. *One-Dimensional Transport with Inflow and Storage (OTIS): A Solute Transport Model for Streams and Rivers*; USGS Water-Resources Investigation Reports 98-4018; U.S. Geological Survey, 1998; http://water.usgs.gov/software/OTIS/ (accessed July 15, 2011).

(27) Allison, J. D.; Brown, D. S.; Novo-Gradac, K. J. *MINTEQA2/PRODEFA2, A Geochemical Assessment Model for Environmental Systems: Version 3.0 User'S Manual, Report*, EPA/600/3-91/021; U.S. Environmental Protection Agency, 1991.

(28) Runkel, R. L.; Kimball, B. A.; McKnight, D. M.; Bencala, K. E. Reactive solute transport in streams: A surface complexation approach for trace metal sorption. *Water Resour. Res.* **1999**, *35*, 3829–3840.

(29) Dzombak, D. A.; Morel, F. M. M. *Surface Complexation Modeling: Hydrous Ferric Oxide*; John Wiley & Sons: New York, 1990.

(30) Kimball, B. A.; Runkel, R. L. Evaluating remediation alternatives for mine drainage, Little Cottonwood Creek, Utah, USA. *Environ. Earth Sci.* **2010**, *60*, 1021–1036.

(31) Box, G. E. P.; Hunter, W. G.; Hunter, J. S. *Statistics for Experimenters*; John Wiley & Sons, 1978.

(32) Müller, B.; Sigg, L. Adsorption of lead(II) on the goethite surface: Voltammetric evaluation of surface complexation parameters. *J. Colloid Interface Sci.* **1992**, *148*, 517–532.

(33) Ali, M. A.; Dzombak, D. A. Effects of simple organic acids on sorption of Cu2+ and Ca2+ on goethite. *Geochim. Cosmochim. Acta* **1996**, *60*, 291–304.

(34) Balistrieri, L. S.; Box, S. E.; Tonkin, J. W. Modeling precipitation and sorption of elements during mixing of river water and porewater in the Coeur d'Alene River basin. *Environ. Sci. Technol.* **2003**, *37*, 4694–4701.

(35) Dixit, S.; Hering, J. G. Comparison of arsenic(V) and arsenic(III) sorption onto iron oxide minerals: Implications for arsenic mobility. *Environ. Sci. Technol.* **2003**, *37*, 4182–4189.

(36) Colorado Department of Public Health and Environment, 2010, Regulation no. 31, The basic standards and methodologies for surface water (5 CCR 1002−31), amended August 9, 2010; p 203, http://www.cdphe.state.co.us/regulations/wqccregs/100231wqccbasicstandardsforsurfacewaternew.pdf (accessed July 15, 2011).