Environmental Toxicology and Chemistry, Vol. 29, No. 11, pp. 2432–2442, 2010

Printed in the USA
DOI: 10.1002/etc.302

DEVELOPMENT OF A NEW TOXIC-UNIT MODEL FOR THE BIOASSESSMENT OF METALS IN STREAMS

TRAVIS S. SCHMIDT,*†‡§ WILLIAM H. CLEMENTS,§ KATHARINE A. MITCHELL,§ STANLEY E. CHURCH,† RICHARD B. WANTY,† DAVID L. FEY,† PHILIP L. VERPLANCK,† and CARMA A. SAN JUAN†
†U.S. Geological Survey Mineral Resources Program, Denver, Colorado 80225, USA
‡U.S. Geological Survey Fort Collins Science Center, Fort Collins, Colorado 80526, USA
§Colorado State University, Fish, Wildlife, and Conservation Biology, Fort Collins, Colorado 80523, USA

(*Submitted* 22 *January* 2010; *Returned for Revision* 15 *March* 2010; *Accepted* 25 *May* 2010)

Abstract—Two toxic-unit models that estimate the toxicity of trace-metal mixtures to benthic communities were compared. The chronic criterion accumulation ratio (CCAR), a modification of biotic ligand model (BLM) outputs for use as a toxic-unit model, accounts for the modifying and competitive influences of major cations (Ca^{2+}, Mg^{2+}, Na^+, K^+, H^+), anions (HCO_3^-, CO_3^{2-},SO_4^{2-}, Cl^-, S^{2-}) and dissolved organic carbon (DOC) in determining the free metal ion available for accumulation on the biotic ligand. The cumulative criterion unit (CCU) model, an empirical statistical model of trace-metal toxicity, considers only the ameliorative properties of Ca^{2+} and Mg^{2+} (hardness) in determining the toxicity of total dissolved trace metals. Differences in the contribution of a metal (e.g., Cu, Cd, Zn) to toxic units as determined by CCAR or CCU were observed and attributed to how each model incorporates the influences of DOC, pH, and alkalinity. Akaike information criteria demonstrate that CCAR is an improved predictor of benthic macroinvertebrate community metrics as compared with CCU. Piecewise models depict great declines (thresholds) in benthic macroinvertebrate communities at CCAR of 1 or more, while negative changes in benthic communities were detected at a CCAR of less than 1. We observed a 7% reduction in total taxa richness and a 43% decrease in Heptageniid abundance between background (CCAR = 0.1) and the threshold of chronic toxicity on the basis of continuous chronic criteria (CCAR = 1). In this first application of the BLM as a toxic-unit model, we found it superior to CCU. Environ. Toxicol. Chem. 2010;29:2432–2442. © 2010 SETAC

Keywords—Trace metals   Aquatic invertebrates   Biomonitoring   Ecotoxicology   Water chemistry

Elevated concentrations of acid and trace metals in surface water draining mineralized and historically mined lithologies are common in Colorado, USA [1,2]. Exposure to elevated concentrations of trace metals adversely affects aquatic populations and communities [1]. Differential sensitivities of individuals to trace metals result in population-level effects that culminate in an assemblage shift from a sensitive to a metal-tolerant community [1,3]. These properties of benthic macroinvertebrate communities make them useful for evaluating the ecological effects of trace-metal pollution in streams.

Trace-metal uptake by aquatic organisms can occur by association with sediment, directly from the water column, or through dietary exposure [4]. However, total recoverable metals (i.e., the sum of dissolved, colloidal, and solid metal that can be liberated via extraction with mineral acid) from a water sample are not good predictors of toxicity to aquatic organisms [5–7]. Current regulatory models treat the total dissolved metal and the dissolved free metal ion as the primary causes of toxicity in aquatic organisms [7,8].

The bioavailability of a trace metal is affected by a suite of constituents found in surface water [6]. The activities of free metal ions are controlled by factors including pH and alkalinity [9]. Interactions with dissolved organic carbon (DOC) and major anions (e.g., HCO_3^-, CO_3^{2-}, Cl^-) decrease the concentrations of trace metals available to cause toxicity [10]. Competition with major cations (e.g., Ca^{2+}, Mg^{2+}) for the sites of toxic action decreases the amount of metal bound to those sites, thus ameliorating toxicity ([5,11–13]; http://www.hydroqual.-com/wr_blm.html; [14]). Because each individual trace metal interacts with these modifying water-quality parameters and sites of toxic action differently to form a variety of metal species, no universal way to quantify the bioavailability and toxicity of trace metals exists [6].

More frequently than not, streams are impaired by a mixture of trace metals at chronic concentrations that act additively to cause toxicity to aquatic organisms [1]. The cumulative criterion unit (CCU) model is a toxic-unit approach that predicts additive toxicity of trace-metal mixtures to aquatic organisms. The CCU relates the total dissolved concentration of a trace metal to the ambient water quality criterion continuous concentration (CCC) for that metal. The criterion for each metal is hardness—adjusted to account for the protective effect of Ca^{2+} and Mg^{2+} using an equation derived from empirical laboratory observations [7]. Incidentally, in these observations it was found that pH and alkalinity co-varied with hardness and, as a result, these hardness adjustment equations indirectly account for the role of pH and alkalinity (pH, alkalinity, and hardness are covariates) on trace-metal toxicity. However, this is an empirical model and not a mechanistic approach to approximating the toxicity of trace metals to aquatic organisms, making it inappropriate for water in which pH, alkalinity, and hardness do not co-vary [8]. More importantly, these correction factors also do not adjust for the role of other aqueous constituents found in surface waters (e.g., DOC, Cl^-, SO_4^{2-}) that also may play an important role in reducing the activity of free metal ions in solution.

The biotic ligand model (BLM) is an algorithm that predicts acute toxicity of dissolved trace metals, such as the free metal

* To whom correspondence may be addressed (tschmidt@usgs.gov).
Published online 9 July 2010 in Wiley Online Library (wileyonlinelibrary.com).

ion, to aquatic organisms (e.g., Cladocera, fish) [11–13]. Coupled geochemical speciation models (CHESS and WHAM V) are used to quantitatively account for the influence of DOC and major cations and anions on dissolved trace-metal bioavailability [10,14]. Empirical data demonstrate that a constant amount of toxicity results from a critical accumulation of a free metal ion on an organism's respiratory surface, which is called the biotic ligand [5,11–13]. Thus, the BLM calculates the fraction of the total dissolved trace metal (free metal ion) in a water sample that is available to accumulate on the biotic ligand. Toxicity is predicted when the ratio of metal accumulated on the biotic ligand exceeds the amount observed to cause mortality to 50% of a population of standard test organisms. Because the BLM can predict the concentration of metal at which 50% mortality occurs to a population of standard test organisms, usually within a factor of 2, the U.S. Environmental Protection Agency (U.S.EPA) has adopted the model to establish site-specific water-quality criteria for Cu [8]. Although BLM models for Cd and Zn (metals of interest in this investigation) as well as for other metals have been developed, these models currently are not employed for regulatory purposes.

Biological assessments such as the Environmental Monitoring Assessment Program, Regional Environmental Monitoring Assessment Program, Wadable Streams Assessment, and the current study, the Central Colorado Assessment Program, are conducted at large spatial scales (e.g., continental United States, Colorado, the central Colorado Mountains) over which climate, vegetation, and geology change from site to site. Commensurate with these changes in landscape characteristics are changes in the concentrations of major cations, anions, and DOC that alter the bioavailability of contaminants in aquatic ecosystems [2,15,16]. Regional-scale biological assessments of trace-metal contamination would benefit from a model that incorporates site-specific variation in aqueous chemistry to more precisely approximate the bioavailable fraction of trace metals to aquatic organisms [17,18].

The BLM is capable of modeling the bioavailable fraction of dissolved trace metals in surface water. However, the BLM was designed to predict acute toxicity (i.e., 48-h concentration of metal at which 50% mortality occurs) to fish and was empirically calibrated to predict toxicity to aquatic invertebrates (i.e., Daphnids) [19]. As a result, some of the underlying mechanisms, although robust, are not specific to the physiology of aquatic invertebrates [19]. For example, the current BLM may not appropriately model Ca^{2+} and Mg^{2+} in competitive interactions with trace metals for the biotic ligand in aquatic invertebrates [19]. The primary assumptions are that all aquatic organisms have a biotic ligand that responds to trace metals in the same general way and that the biotic ligand is the primary pathway of toxicity. However, the BLM has not been tested to determine whether the criteria set by the model are protective of indigenous organisms under field conditions.

Applied as intended, the BLM is not especially useful for bioassessment. The primary output of the BLM (concentration of metal at which 50% mortality occurs) is not ecologically meaningful, because it does not describe possible consequences to higher levels of biological organization (e.g., populations, communities, ecosystems) [20]. Neither does the model predict the effects of metal mixtures on aquatic communities, a situation more common in areas influenced by acid rock drainage [1,2]. If the BLM is to be employed to set site-specific water-quality criteria protective of aquatic communities, it should be capable of predicting the responses of natural populations and communities in streams [21].

We have developed a method that uses the BLM for the purpose of ecological assessment of trace-metal pollution in natural systems. We developed a toxic-unit model of additive trace-metal toxicity derived from BLM outputs and compared it with another toxic-unit model, CCU. This new model, the chronic criterion accumulation ratio (CCAR), is derived from BLM outputs, thereby incorporating current theory about the interactions between aqueous constituents (i.e., hardness, DOC, pH) that affect trace-metal toxicity and accumulation of bioavailable trace metals on the respiratory surface of aquatic organisms. In contrast, the CCU accounts only for the influence of hardness on trace-metal toxicity through the use of empirical equations derived from single-species toxicity tests.

The primary objective of this research is to explore the use of the BLM as a bioassessment tool to predict responses of benthic macroinvertebrate communities to trace metals in streams throughout Colorado. We make comparisons between this new BLM-derived estimate of trace-metal toxicity and the model of additive toxicity on the basis of hardness-adjusted chronic criterion values. Cadmium, Cu, and Zn are the metals of interest in this investigation because they have been identified as compounds of concern for this region by previous researchers, they are common to the mineralogy of the region, and they have BLM models developed and commonly available [1,2,13]

MATERIALS AND METHODS

Development of toxic-unit models

Because most trace-metal–contaminated streams are influenced by a mixture of metals at chronic concentrations, a measure of chronic toxicity resulting from metal mixtures was necessary. Water-quality criteria for individual metals represent concentrations that, when exceeded, likely harm aquatic organisms. Because criterion values are established for individual metals, alternative models are necessary to estimate toxic effects of trace-metal mixtures. Although most research investigating the toxicity of trace-metal mixtures has focused on acute effects, previous studies have shown additive effects at chronic concentrations [1,12,19]. The CCU was used to evaluate toxicity resulting from trace-metal mixtures and assumes that interactions among trace metals are additive. The CCU is defined as the ratio of the measured trace-metal concentration to the U.S. EPA hardness-adjusted chronic criterion value, summed for each metal (Cd, Cu, Zn) at a location [7]. The cumulative criterion unit is calculated as:

$$\mathrm{CCU} = \sum_i \frac{m_i}{c_i} \tag{1}$$

where m_i is the total dissolved trace-metal concentration and c_i is the hardness-adjusted continuous chronic criterion (CCC) value for the *i*th metal. Because water hardness affects toxicity and bioavailability of some trace metals, criterion values for Cd, Cu, and Zn were modified to account for variation in water hardness among streams [7]. For example, at a water hardness of 100 mg/L ($CaCO_3$), criterion values for these three trace metals would be 0.25, 9.0, and 120 μg/L, respectively. The CCC is developed by averaging toxicity test data across species and genera to determine a concentration of trace metal that will be protective of 95% of the species at a specific site. Therefore, a CCU value of 1.0 or less represents a mixture of metal concentrations that should be protective of aquatic communities. This model is a common approach for assessing toxicity caused by metal mixtures and will be used to evaluate the BLM predictions.

Because the BLM predicts acute toxicity resulting from individual trace metals, it must be modified to account for metal mixtures and for comparison with the CCU model. The chronic criterion accumulation ratio (CCAR) is a procedure that modifies BLM outputs for use as a toxic-unit model similar to the CCU model and assumes additive toxicity of trace-metal mixtures on the basis of BLM-predicted outputs. The CCAR is defined as the ratio of the BLM-calculated accumulated free metal ion on the biotic ligand to that accumulated on the biotic ligand in water at the U.S. EPA criterion value, summed for trace metals of interest at a location. The CCAR is calculated as follows:

$$\mathrm{CCAR} = \sum_i \frac{\text{BLM calculated site specific [gill metal]}}{\text{BLM calculated [gill metal] at CCC}} \quad (2)$$

where BLM calculated site-specific [gill metal] and BLM calculated [gill metal] at CCC are measurements developed from BLM outputs. The BLM calculated [gill metal] is the BLM predicted accumulation of the *i*th trace metal on the biotic ligand (gill surface), calculated by running the BLM in speciation mode, using site-specific water quality parameters (temperature, pH, DOC, alkalinity, Ca^{2+}, Mg^{2+}, Na^+, K, SO_4^{2-}, S^{2-},Cl^-). The BLM calculated [gill metal] at CCC is the BLM predicted accumulation of the *i*th trace metal on the gill surface, calculated using the "normalization chemistry" water-quality parameters from Table 1 of the U.S. Environmental Protection Agency water quality criterion document [8] and the *i*th metal CCC [7]. A CCAR of 1.0 or less represents a mixture of metal concentrations at or below CCC, accounting for the modifying effects of several water-quality parameters known to alter trace-metal bioavailability, and protective of an aquatic community. We used CCAR to predict toxicity to benthic communities, individual species of which have differential sensitivity to trace metals. By using the CCC, a value derived to protect aquatic communities, rather than a species-specific response point such as the amount observed to cause mortality to 50% of a population of standard test organisms for Daphnids or fathead minnows, we can predict toxicity more generally to the entire benthic community. Specifics on model constants and assumptions can be found in HydroQual [13].

Study area and sampling strategy

The study area is central Colorado from Wyoming to New Mexico, USA, an area of approximately 54,000 km^2 that includes most of the Rocky Mountains in Colorado and represents approximately 20% of the land area of Colorado (Fig. 1). This area includes a geographic feature called the Colorado Mineral Belt that has been exploited for the past 150 years for its mineral resources. The sample sites in the current study are at high altitude, ranging from 2,330 to 3,550 meters above sea level. The climate of the study area is temperate continental, with generally more than 50 cm precipitation per year, especially at higher altitudes. Much of this precipitation occurs as snow in winter or as rain, primarily between June and August. Vegetation ranges from deciduous cover at lower altitudes and in riparian zones, to conifer forests, and at the highest altitudes, open tundra. Soils within the study area are thin (rarely greater than 10 cm) to nonexistent, the latter occurring in areas dominated by bedrock outcrops. Thicker (up to a meter or more) immature soils, as well as unconsolidated overburden, occur intermixed at lower elevations and along streams.

Small catchments (first to third order) predominantly underlain by a single rock-type categorized on the basis of lithology were targeted for sampling. The purpose of this sampling strategy was to target a large variety of water-quality conditions resulting from interaction with the underlying rocks to test the BLM and develop lithologic-specific geochemical baselines [2,22]. This approach uses geological principles to identify locations expected to have low metal concentrations and locations in which metal concentrations are expected to be high. Each catchment was characterized as to the presence or absence of geological processes that influence the acidity and trace metals found in catchment bedrock (i.e., hydrothermal alteration and ore deposit formations) and the presence or absence of mining. Further details about how hydrothermal alteration and ore deposits were used for site selection can be found in Schmidt et al. [23].

Geochemical and benthic macroinvertebrate samples were collected from 153 catchments during base-flow conditions in the summers of 2003 ($n = 20$), 2004 ($n = 41$), 2005 ($n = 38$), 2006 ($n = 31$), and 2007 ($n = 23$) (Fig. 1). All geochemical and benthic macroinvertebrate samples were either collected simultaneously, or in a few cases, within a 10-d period of each other. Geographic Information Systems (ArcGIS 9.2) were used to delineate catchments for sampling, and digital elevation models (30×30 m) were used to define catchment boundaries, area, slope, and relief ratio [24]. The Colorado Vegetation Model (http://warnercnr.colostate.edu/~davet/cvm.html; 30 m × 30 m

Table 1. Benthic macroinvertebrate metrics selected on the basis of performance in past regional-scale assessment of the effects of metals on stream communities

Metrics	Clements et al. [11][a]	Crane et al. [17][b]	Griffith et al. [33][c]	Hirst et al. [18][d]	Malmqvist et al. [32][e]
Richness metrics					
Total taxa richness	X		X	X	X
EPT[f] richness	X	X	X		X
Abundance metrics					
Total abundance	X		X	X	
Mayfly abundance	X				
Heptageniidae abundance	X				
Functional metrics					
Scrapper abundance	X				
Predator abundance	X				

[a] Metrics distinguished different levels of metals contamination based on cumulative criterion unit (CCU), Colorado Rocky Mountains, USA.
[b] Metrics used to assess European Water Framework Directive Quality Standards for dissolved metals in England and Wales, United Kingdom.
[c] Metrics that detect differences between sites that were above or below environmental quality standards for water or sediment, Colorado Rocky Mountains, USA.
[d] Metrics were significantly correlated with CCU metric in Wales and Cornwall, United Kingdom.
[e] Metrics detected differences in streams contaminated with metals in Dalarna Provence, Sweden.
[f] EPT= Ephemeroptera + Plecoptera + Trichoptera.

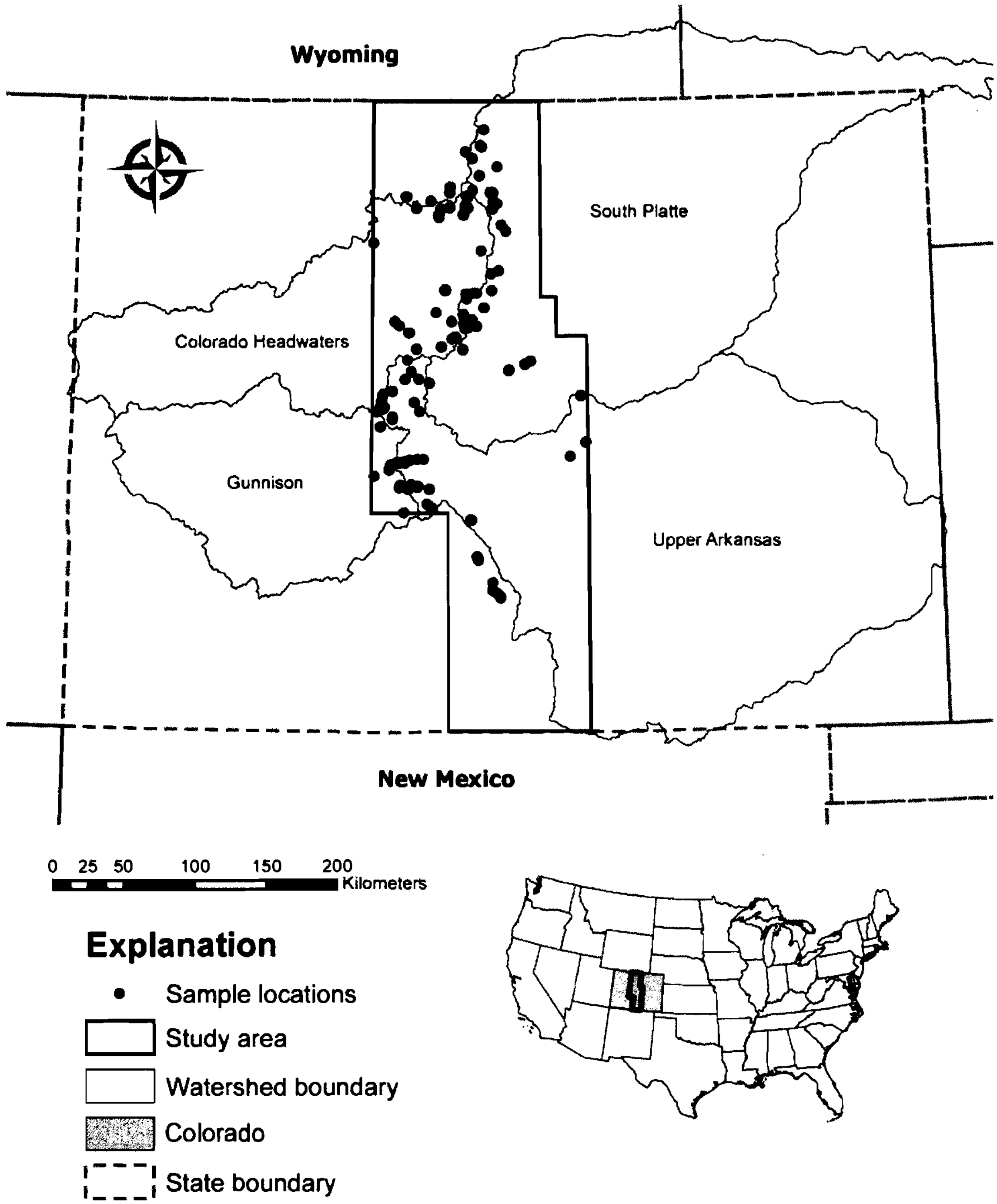

Fig. 1. Map of the United States showing the central Colorado study area in the Rocky Mountains from New Mexico to Wyoming.

resolution) was used to eliminate areas of development (i.e., agricultural, residential, and commercial development) to make sure we targeted catchments that were not influenced by anthropogenic factors other than mining.

Physicochemical parameters

Current velocity was measured with the U.S. Geological Survey Price pygmy current meter and depth were measured across the stream channel at 15 to 25 intervals, depending on stream width. Stream discharge (f^3/s) was calculated using the continuity equation. Stream substrate size measurements and densiometer (Forest Densiometers, Model A) readings of canopy cover were collected from each location where benthic samples were collected from 2003 to 2005. After the third year of the study, we screened the data (i.e., exploratory statistical analysis) to determine whether local habitat variables were important determinants of benthic macroinvertebrate community structure. Stream substrate size (D10, D50, D95, frequency of all size classes from 22.5 to 180 mm) and canopy cover contributed little to no descriptive power in statistical models predicting biological responses (Bray-Curtis similarity); these two habitat parameters were not collected in the final two years of the study.

Water samples in 2004 to 2007 were collected using methods described in Wilde et al. [25] to meet the requirements of the

BLM [13,25]. Routine water-quality parameters (temperature, conductivity, and pH) were measured in the field using a Horiba D-24 combination meter [26]. Meters were calibrated at the beginning of each day with certified standards and checked periodically throughout the day. All water samples processed in the laboratory were filtered through an Acrodisc Premium 25-mm Syringe Filter with 0.45-μm nylon membrane at the site and stored at 4°C until analyzed. Water samples for DOC were filtered through a 0.70-μm glass-fiber filter, acidified with concentrated hydrochloric acid (12 molar) to a pH of less than 2, and stored in baked amber glass bottles. A Shimadzu TOC-5000A total organic carbon analyzer was used to measure DOC. Dissolved trace-metal samples were acidified with concentrated nitric acid (13 molar) to a pH less than 2 and stored in polyethylene bottles. Samples for anions were collected and stored in polyethylene bottles. Water analyses were conducted at the analytical laboratories of the U.S. Geological Survey Geologic Discipline Laboratory in Denver, Colorado. Concentrations of major cations (Na^+, K^+, Mg^{2+}, and Ca^{2+}) were analyzed by inductively coupled plasma-atomic emission spectrometry, whereas trace-metal concentrations (Cd, Cu, Zn) were analyzed by inductively coupled plasma-mass spectrometry, and major anions (Cl^-, F^-, NO_3^-, SO_4^{2-}) were measured by ion chromatography [26]. High concentrations of SO_4^{2-} (>25 mg/L) were determined by inductively coupled plasma-atomic emission spectrometry, and alkalinity was determined by titration [26]. For metal concentrations below detection, half the detection limit was substituted as the value.

Analysis methods for 2003 differed in that major cations and trace metals were analyzed by flame (Zn) and furnace atomic absorption (Cd and Cu) spectrophotometry (PerkinElmer model 372). This analysis was conducted at the Colorado State University Fish, Wildlife, and Conservation Biology Department, Fort Collins, Colorado, USA. The detection limits were different between sampling periods as follows; Cd (0.01 μg/L in 2003 vs 0.02 μg/L in 2004–2007), Cu (0.01 μg/L in 2003 vs 0.5 μg/L in 2004–2007), and Zn (5 μg/L in 2003 vs 0.5 μg/L in 2004–2007). These detection limits are equal to or lower than those commonly reported for field assessments in the literature, and well below any previously reported toxic concentrations for these metals. These differences in minimum reporting limits between 2003 and all other years are unlikely to be meaningful.

Benthic macroinvertebrate sampling

Five replicate benthic samples ($n = 5$) were collected using a 0.1-m^2 Hess sampler (350-μm mesh net) from shallow riffle areas (<0.5 m). Representative sample localities were selected on the basis of the following criteria: location was a riffle or run habitat unit, depth was 0.10 to 0.25 (m), and substrate was representative of the stream reach. Overlying substrate was scrubbed of algae and diatoms, and inorganic debris was removed. All individual substrate particles larger than 22.6 mm were removed from the Hess sampler and measured along the intermediate axis. Underlying substrate was disturbed to a depth of approximately 10 cm, and the remaining material was sieved using a 350-μm mesh sieve. All organisms retained were preserved in 80% ethanol in the field.

In the laboratory, samples were processed to remove debris and subsampled until 300 organisms (±10%) were removed from the sample, using methods described in Moulton et al. [27]. Invertebrates were identified to the lowest practical taxonomic level (genus or species for most taxa; subfamily for chironomids) [28,29]. The Invertebrate Data Analysis System (U.S. Geological Survey, USA) version 4.2.10 was used to resolve taxonomic ambiguities and reduce the influence of rare taxa on this large-scale regional environmental assessment [30]. Taxa that were not found to occur in at least 20 sites were dropped from analysis, to reduce the influence of rare taxa on study results. Ambiguities in community data sets occur when closely related specimens are identified at different levels of taxonomic resolution. This usually occurs because the variation in life history of closely related species results in a wide range of individual maturity. Characteristics used to separate species or genera are developed from mature specimens and may not be present in earlier instars or damaged individuals. As a result, a sample may contain a group of individuals from the same family, but not all can be identified to the genus level; the resulting taxa list may show some identified to genera and others only identified to family. When characteristics needed to identify an organism to a finer level (for example, species level) are not present, assumptions about their identity beyond the coarser level (for example, genus level) lead to ambiguities. Including ambiguous taxa in a data set can inflate richness or other measures of community structure. Ambiguities resulting from differing levels of identification were resolved by distributing individuals identified at coarser levels to finer levels dependent on their abundance.

Means of the five replicate benthic samples, once processed in the Invertebrate Data Analysis System as described, were used to calculate benthic macroinvertebrate community metrics (Table 1) [29]. These metrics were selected based on a literature review of large-scale (at least 50 locations investigated in multiple watersheds) bioassessment studies evaluating the effects of metals on benthic macroinvertebrates [1,17,18,31,32]. For a comprehensive analysis of stream communities, we sought to include measures of community richness, abundance, and function. We considered all benthic macroinvertebrate metrics that were significantly related to metals contamination from these past studies and refined the list to include those metrics used in multiple studies or distinguished between multiple levels of metal contamination in the Clements et al. study [1].

Data analysis

All statistical analyses were conducted in R version 2.7.2 ([33]; http://www.R-project.org). Scatter plots of the data suggested nonlinear and possibly threshold responses by benthic macroinvertebrate communities. Piecewise linear regressions [34] fit two linked line segments connected at a threshold where an abrupt change in response required a different slope to fit the regression. The threshold and 95% confidence intervals were estimated using a bootstrap method resampling the raw data 1,000 times [34]. Akaike information criteria (AIC) values were calculated to determine which of the two competing models (CCAR or CCU) had the highest probability of being the best model [35]. We used a version of AIC (AIC_c) that corrects for a small sample size. The AIC_c was standardized by subtracting the minimum AIC_c score from each of the candidate model AIC_c values to derive Δ_I and facilitated the ranking of candidate models [35]. Akaike weights (ω_i) were calculated to determine the probability of a model being the best model among those in the candidate set. For thoroughness, Spearman correlations between the two toxic unit models and habitat variables were evaluated to determine whether the difference in the predictive nature between these two toxic unit models was attributable to collinearity with habitat variables.

Unlike previous research conducted in this region [1,32], the current study made an effort to target streams with relatively low contaminations of trace metal. We were interested in determining whether effects could be detected at these lower metal concentrations. First, data were binned into two categories of toxic-unit ranges, sample locations with CCAR at least 0.1 and sample locations with CCAR greater than 0.1 but less than or equal to 1.0. These values were selected because previous analysis suggests that CCAR = 0.1 is the average background value for all catchments in the study area not influenced by lithologies with acid-generating capacity or that release metals into streams. The upper range (CCAR = 1.0) was selected because this value is expected to protect aquatic communities [21,23]. This approach eliminates leverage on the statistical model caused by obvious declines in benthic macroinvertebrate communities at very high metal concentrations. Because nonlinearity and heterogeneity of the response variables were observed in the scatter plots and sample size was unbalanced ($n = 34$ vs 74), a nonparametric Mann-Whitney *U* test ($p \leq 0.05$) was employed to quantify differences between categories [36]. Percent differences in mean metric values also were calculated.

RESULTS

Physicochemical characteristics

Habitat characteristics of targeted catchments were typical of small to mid-sized headwater streams (i.e., 1st to 3rd order) of the Rocky Mountains and indicated that anthropogenic influences other than mining (i.e., mean area as agriculture = 0% ± 1 SD, mean area developed; urban and commercial development = 0% ± 2 SD) had little direct effect on water quality and benthic macroinvertebrate communities (Table 2). The large variation observed in some local (e.g., D50, discharge, and percent cover) and landscape-level (e.g., percent forest area) parameters were expected because we sampled catchments underlain by different lithologies to capture a wide range of both chemical and physical differences of streams within the study area (Table 2). Spearman correlations (Table 3) between the two toxic-unit models and habitat variables were weak ($r < 0.6$), and no differences were found in the pattern of intercorrelations, suggesting that differences in the performance of the two toxic-unit models were not attributable to collinearity with habitat variables.

Field sites in the current study showed a broad range of variation in water quality; however, these values were within the range of water-quality parameters used to develop the BLM (Table 4) [13]. Exceptions included temperature (3–18°C), pH (3.5–8.5 pH), Cl^- (0.04–9.5), and alkalinity (0–141 mg/L), which were occasionally less than the specified limits of the BLM.

Table 2. Range and mean ± standard deviation (SD) for habitat parameter measured in study catchments

Parameter	Range	Mean ± SD
D_{50} (mm)[a,b]	22.6–90	38 ± 14
CFS[c]	0.2–137	11 ± 15
Cover (%)[b,d]	1–96	58 ± 29
Alpine (%)[e]	0–97	36 ± 28
Forest (%)	2–95	44 ± 36
Agriculture (%)[f]	0–1	0 ± 0.0
Developed (%)[g]	0–2	0 ± 0.0
Site elevation (m)	2,329–3,535	2,975 ± 287
Watershed area (km²)	2–480	36 ± 58

[a] Median particle size in millimeters.
[b] Not measured in years 2006 to 2007.
[c] CFS = cubic feet per second.
[d] Percent area obscuring the sky.
[e] Sum of area as bare ground and ice/snow.
[f] Sum of area as pasture and hay.
[g] Sum of area as residential or commercial development.

Table 3. Spearman rank correlations between habitat factors and toxic-unit models[a]

Variable	CCU[b]	CCAR[c]
CFS[d]	*0.39*	*0.39*
Alpine (%)[e]	*0.39*	0.44
Forest (%)	*−0.39*	*−0.46*
Agriculture (%)[f]	0.11	0.04
Developed (%)[g]	0.14	0.15
Site elevation (m)	*0.25*	*0.36*
Basin area (km²)	*0.25*	0.19

[a] Italic correlation coefficients were significant at $p \leq 0.005$ to correct for experimentalwise error (i.e., Bonferroni's adjustment).
[b] CCU = cumulative criterion unit.
[c] CCAR = chronic criterion accumulation ratio.
[d] CFS = cubic feet per second.
[e] Sum of area as bare ground and ice/snow.
[f] Sum of area as pasture and hay.
[g] Sum of area as residential or commercial development.

Comparison of models of trace-metal toxicity

Calculations of CCU in the current study differ from previous investigations [1] in that toxicity was estimated using total dissolved trace-metal concentrations instead of total metal concentrations. However, dissolved metal concentrations were only marginally lower than total metal concentrations in the current study. Other studies in this region included Al, Fe, Mn, or Pb in their calculations of CCU [1,32]. These metals were not included in our toxic-unit calculations because BLM models are not available for these metals, and therefore no comparisons could be made between toxic-unit models. Cadmium, Cu, and Zn cause toxicity by disrupting similar physiological processes, unlike the other metals, and these metals were observed at low concentrations as compared with Cd, Cu, and Zn. Of the four metals not included in our CCU calculation, Al and Fe exceeded CCC (87 and 1,000 μg/L, respectively) at two (Al) and one (Fe) locations. By incorporating these four metals into our CCU calculation, only 12 additional sites exceed CCU = 1. This demonstrates that few sites (8%) were falsely classified as below the threshold thought benign to aquatic life, because we did not include all the metals measured in water in our index of trace-metal toxicity [1]. Inclusion of these metals in regression models did not improve model fit.

Chronic criterion accumulation ratio values for the sum of Cd, Cu, and Zn ranged from 0.02 to 268, whereas CCU values at these stations ranged from 0.05 to 125. A direct comparison of the two models of trace-metal toxicity shows that CCAR underpredicts toxicity, especially at low metal concentrations (Fig. 2A). Although CCAR calculates the bioavailability of free metal ions and CCU calculates the bioavailability of total dissolved metals, the difference in the magnitude of concentrations does not explain the underprediction, because each toxic unit is normalized and therefore unitless. However, by including DOC in the calculation of CCAR, the amount of free metal ion available to bind to the biotic ligand is decreased, especially at low metal concentrations. Chronic criterion accumulation ratio overpredicted toxicity relative to CCU at 20 sites. In 17 of these cases, alkalinity was less than 8 mg/L, and in all of these cases the total dissolved metal or free metal ion was relatively high compared with all other cases. In 11 other cases,

Table 4. Range, median ± standard deviation (SD), and biotic ligand model design range for chemical parameters measured in the current study

Parameter	Range	Median ± SD	BLM[a] design range [13]
Temperature (°C)	3–18	9.6 ± 2.75	10–25
pH (Standard units)	3.5–8.5	6.9 ± 1.0	4.9–9.2
Dissolved organic carbon (mg/L)	0.3–8.1	1.4 ± 1.10	0.05–29.65
Humic acid content (%)		10[b]	10–60
Hardness (mg/L)	5–163	41 ± 31	NA[c]
Ca^{2+} (mg/L)	1.5–45.4	12.0 ± 8.4	0.204–120.24
Mg^{2+} (mg/L)	0.2–16	2.2 ± 2.8	0.024–51.9
Na^{+} (mg/L)	0.2–36	1.5 ± 3.1	0.16–236.9
K^{+} (mg/L)	0.15–4.2	0.55 ± 0.55	0.039–156
SO_4^{2-} (mg/L)	0.7–167	7.8 ± 28.8	0.096–278.4
Cl^{-} (mg/L)	0.04–9.5	1.20 ± 1.42	0.32–279.72
Alkalinity (mg/L)	0–141	25 ± 26	1.99–360
DIC (*mmol*/L)[d]	Estimated	Estimated	0.056–44.92
S^{2-} (mg/L)	NM[e]	NM[e]	0–0
Cd^{2+} (μg/L)	0.01[f]–7.92	0.01[f] ± 1.07	NL[g]
Cu^{2+} (μg/L)	0.15–935	0.50 ± 13.6	NL
Zn^{2+} (μg/L)	0.25[h]–1940	3.40 ± 254.81	NL

[a] BLM = Biotic ligand model.
[b] Not measured but default value was used as recommended, 10% [13].
[c] Not applicable.
[d] Dissolved inorganic carbon = BLM estimates DIC from measured values of alkalinity and pH [13].
[e] Not measured but default value used as recommended, 1 · E − 10 [13].
[f] Detection limit.
[g] No limits are set for trace-metal concentrations.
[h] Half detection limit.

alkalinity was less than 8 mg/L; however, in these cases CCU overpredicted toxicity relative to CCAR, and they all had relatively low metal concentrations.

Each metal's contribution to a toxic-unit model changed depending on the concentration of the metal and toxic-unit model considered (Fig. 2B, C). The BLM incorporation of DOC resulted in a lower contribution of Cu at toxic-unit values below 10, as compared with CCU. At toxic units of 10 or greater, Cu was found to dominate toxic-unit values, the result of high concentrations of Cu, a decrease in the capacity for DOC to bind Cu, and the inability for hardness to competitively interact with the biotic ligand to ameliorate toxicity. Most sites with toxic-unit values of 10 or greater were found to have low alkalinities and pH and consistently fell below the regression line (Figs. 3–5). These sample locations were also influenced by substantially higher concentrations of Al and Fe.

Community-level responses to metals

Piecewise linear regression analyses using AIC for model selection found that CCAR was the most likely predictor of the benthic macroinvertebrate metrics (Figs. 3–5). The cumulative criterion unit received no weight as the top model in most cases, with the highest observed likelihood = 0.07. The threshold for all metrics evaluated exceeded CCAR = 1: richness = 1.06, EPT (Ephemeroptera + Plecoptera + Trichoptera) richness = 1.11, total abundance = 1.47, Ephemeroptera abundance = 1.21, Heptageniidae abundance = 1.16, predator abundance = 1.19, and scrapper abundance = 1.16 (Figs. 3–5). However, in almost every case the 95% confidence interval ranged well below CCAR = 1: richness = (0.17–1.25), EPT richness = (0.39–1.68), Ephemeroptera abundance = (0.76–2.07), Heptageniidae abundance = (0.69–1.97), predator abundance = (0.89–4), and scrapper abundance = (0.35–1.49), the lone exception being total abundance = 1.47 (1.06–5.52), (Figs. 3–5).

Visual inspection of Figures 3 to 5 indicates that trace-metal concentrations (characterized as CCAR) caused various effects in benthic macroinvertebrate communities from central Colorado mountain streams. Slopes for all piecewise linear models were negative, suggesting negative effects occurred below the threshold of chronic toxicity. This is also supported by the confidence intervals ranging well below CCAR = 1 (Figs. 3–5). Results of a Mann-Whitney U test showed significant differences in mean metric values between background sites and CCAR = 1.0, the theoretical threshold of toxicity (Table 5). Significant differences were observed in total taxa and EPT richness (−7 and −6% mean differences, respectively), Ephemeroptera and Heptageniidae abundance (−33 and −43% mean differences, respectively), and predator abundance (−26% mean difference). The number of samples included in each category was different, $n = 34$ versus $n = 74$, and although the ranges of benthic macroinvertebrate metric values were similar, a proportionately larger number of samples were observed to have lower benthic macroinvertebrate metric values at CCAR = 1.0 as compared with background (Figs. 3–5).

DISCUSSION

The intent of this work was to develop and evaluate the BLM as a bioassessment tool capable of predicting benthic macroinvertebrate community responses to trace-metal mixtures. This is an important step in improving our ability to link field-based community and population responses to metal toxicity while including the modifying effects of water quality in the determination of trace-metal bioavailability [17,18]. However, the BLM was developed to predict toxicity to standard test organisms and has not been evaluated to determine whether it can predict the response of communities to trace metals in surface water [9]. We developed a toxic-unit model that uses BLM-derived outputs (CCAR) to predict responses of benthic macroinvertebrate communities to mixtures of trace metals and evaluated its performance relative to the CCU. This evaluation was conducted as part of a regional-scale environmental assessment of the effects of geology on the environment in Colorado, which provided a diversity of lithologies and physicochemical conditions to test this new model of trace-metal toxicity.

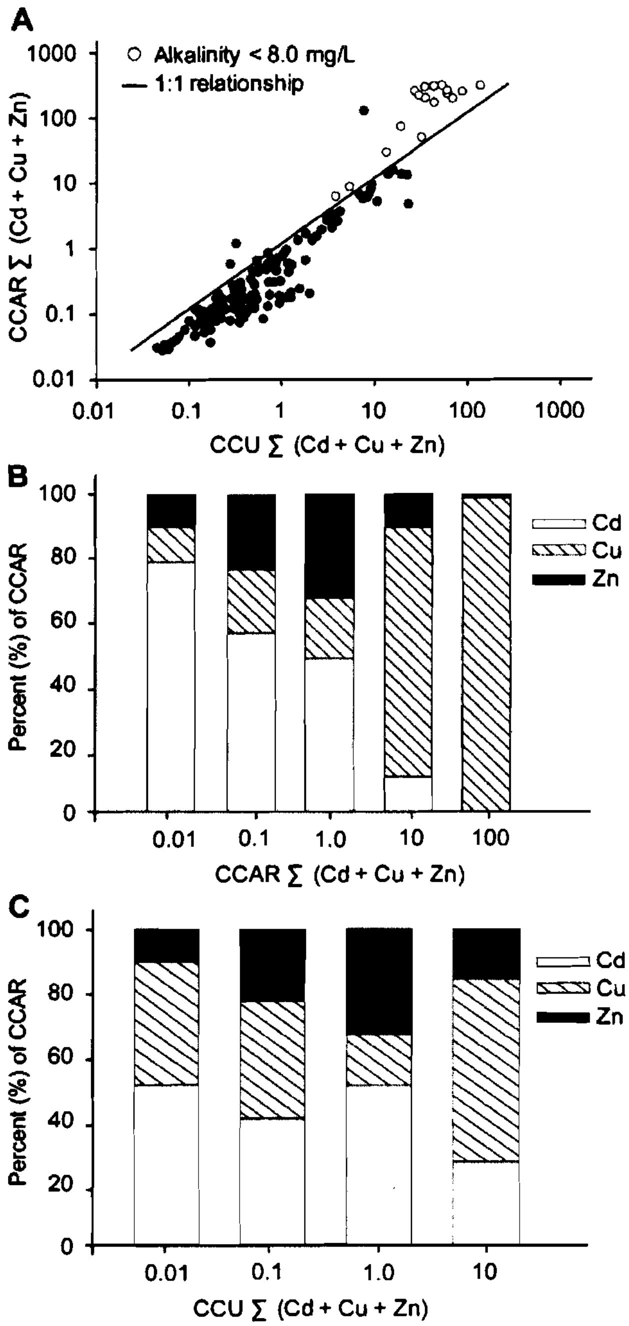

Fig. 2. Graphs depict the relationship between $\log_{10}$ (CCU) (cumulative criterion unit) and $\log_{10}$(CCAR) (chronic criterion accumulation ratio) (**A**), and the relative contribution of each metal (Cd, Cu, Zn) in terms of toxic units to CCAR (**B**) and CCU (**C**). Filled circles, CCAR versus CCU.

Streams in the Colorado Rockies are generally oligotrophic, with low concentrations of dissolved solutes. We observed this to be true of the streams we sampled, in which most physicochemical characteristics were within the range of those used to develop the BLM. However, a number of water-quality parameters (e.g., pH, alkalinity, Cl^-) were found to be on the lower end of the design range. This condition generally increases the bioavailability of trace metals as compared with water of high ionic strength and DOC concentrations [5,14]. However, CCAR, which accounts for DOC, measured a decrease in trace metal available as compared with CCU in streams with relatively low metal concentrations. In contrast, at locations where low pH and high trace metal concentrations co-occurred, CCAR predicted higher toxic-unit values as compared with CCU. This increase in trace-metal bioavailability was substantial and demonstrates that low pH can overwhelmingly increase trace-metal toxicity despite the influence of other water-quality parameters [8]. These results highlight the need to incorporate a suite of water-quality parameters that can affect trace-metal bioavailability in field studies [16,17].

Chronic criterion accumulation ratio is far superior to CCU as a predictor of benthic macroinvertebrate community responses to metal mixtures. We anticipated that CCAR would outperform CCU because CCAR incorporates a mechanistic understanding of the chemical processes that control trace metal bioavailability to aquatic organisms and accounts for physiological processes. However, the BLM was recalibrated to model invertebrate responses to acute concentrations of free metal ion; mechanisms of toxicity to chronic exposure to trace-metal mixtures may be different [19]. Under conditions of chronic exposure, benthic macroinvertebrates regulate, detoxify, and eliminate metals, processes not modeled by the BLM [37,38]. Nor does the BLM recognize the accumulation or regulation of metals through the diet [38]. Our empirical statistical analysis suggests that CCAR is a better predictor of benthic macroinvertebrate community responses as compared with CCU; however, this approach should be improved on to make it more mechanistic.

We identified that benthic macroinvertebrate community metrics exhibited a threshold response to trace-metal mixtures. The 95% confidence interval for this threshold included CCAR = 1 in every metric except for total taxa abundance. This finding could be interpreted as evidence that the CCC or the BLM are protective of aquatic life; however, this is not the case. The piecewise linear models identified a threshold at which higher rates of decline in benthic macroinvertebrate community metrics were observed at CCAR near 1, whereas lower rates of decline were observed at CCAR < 1. We conclude that profound changes in aquatic communities were observable near CCAR = 1, whereas measurable but highly variable responses were observed at toxic-unit values below 1.

Previous studies report significant reductions in community richness and abundances of sensitive taxa at 2 CCUs or greater [1,39]. Other researchers suggest that benthic macroinvertebrate communities responded negatively to trace-metal mixtures only once concentrations exceeded the threshold of chronic toxicity, CCU = 1.0 [32]. The results of our study corroborate earlier findings that trace metals negatively affect benthic macroinvertebrate communities at or near water-quality criterion. Novel to our study are data that suggest losses in benthic macroinvertebrate richness, abundance, and function occurred at concentrations below the CCC. These previous studies included other metals (Al, Fe, Pb) into their calculation of CCU, or utilized total metals [1,39] or total dissolved metals [32], which would increase the toxic-unit values at all sites, forcing the observed declines in benthic macroinvertebrate communities to occur at higher toxic-unit values. Our direct comparison of the predictive capacity of free metal ion versus the total dissolved metal fractions suggests that the free metal ion is a better predictor of biological responses. Had earlier investigations considered free metal ion concentrations, likely their observations of adverse effects would occur at lower toxic-unit values.

Our observed changes in benthic macroinvertebrate communities below CCC does not mean that water quality standards

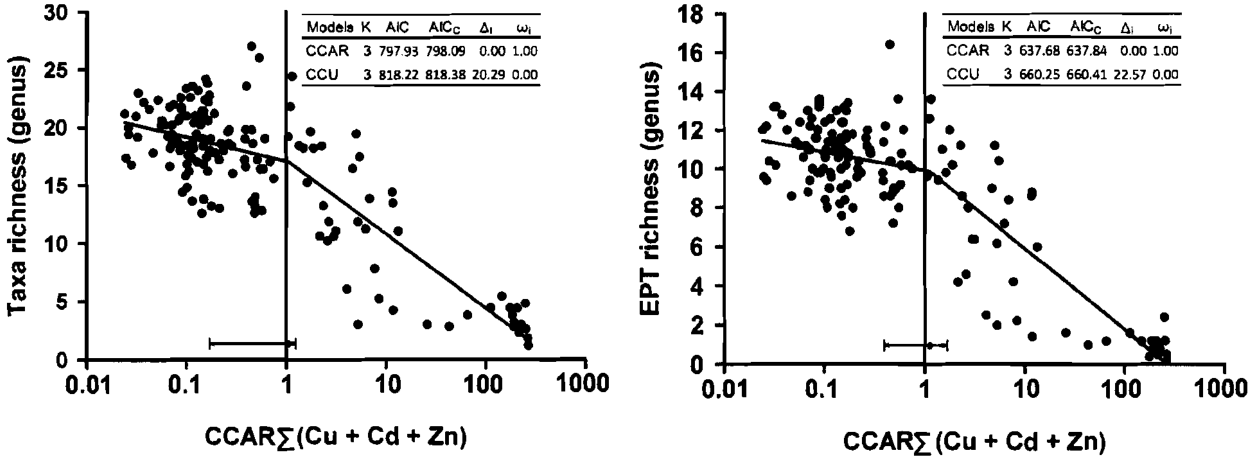

Fig. 3. Piecewise linear regressions showing the relationship between benthic macroinvertebrate community richness metrics and $\log_{10}$(CCAR) (chronic criterion accumulation ratio). Summaries of model selection analysis are depicted in the table showing Akaike information criteria values. Vertical lines indicate the limit above which adverse effects in aquatic communities are expected. Horizontal segmented lines are the piecewise linear association between richness metrics and CCAR. The 95% confidence interval for piecewise threshold is depicted by horizontal arrows and brackets, and the threshold is identified as the black circle within the confidence intervals. AIC = Akaike information criteria; AIC_c = second-order correction of AIC to account for small ratio of model parameters (K) to observations (n); $\Delta_i = AIC_c$ standardized by subtracting the minimum AIC_c score from each of the candidate models; ω_i = Akaike weight or the probability the model is the best model among the candidate set; CCU = cumulative criterion unit; EPT = Ephemeroptera + Plecoptera + Trichoptera.

or the BLM are not protective of aquatic life in Rocky Mountain streams. Water quality standards were established to protect 95% of genera [21] in a water body, not just benthic macroinvertebrates. Because we did not measure responses in all the necessary phyla, classes, and families (e.g., fish, algae, crustacean) specified in the guidelines for deriving water quality criteria [21], we did not test the suitability of these standards to protect aquatic life. Also, water quality standards were established to protect aquatic communities in receiving waters of the discharge of a single pollutant, not nonpoint sources of complex trace-metal mixtures. However, we suggest that if water quality criteria are used as a benchmark of successful restoration of complex acid rock drainage, benthic macroinvertebrate communities exposed to trace-metal mixtures at these concentrations may not recover to levels found in streams with lower concentrations of trace-metal mixtures.

The observation that population-level abundance metrics suffer greater losses than richness metrics when exposed to similar levels of metal mixtures is not a novel finding. Previous studies, both field and laboratory based, have documented that sensitive mayflies and stoneflies experience population declines at much greater rates than observed in generic-level total taxa richness [1,40]. Furthermore, single species toxicity tests using a sensitive mayfly of the genus *Rhithrogena* have shown that mature nymphs can tolerate very high metal concentrations and survive an acute exposure, whereas populations of less mature individuals suffer great loses at lower metal concentrations [39–41].

The current study confirms that richness metrics are not the most sensitive indicators of the effects of trace-metal mixtures on aquatic communities in streams. However, we observed great declines in the abundance of metal-sensitive benthic macroinvertebrate populations in response to levels of trace-metal mixtures never reported before. Such declines in abundance cause disruptions of in-stream ecosystem function such as the processing of detritus, secondary production of invertebrates, and flow of energy into food webs [42]. This disruption of energy flow into food webs is not limited to aquatic food webs, because disturbance-induced declines in abundance of

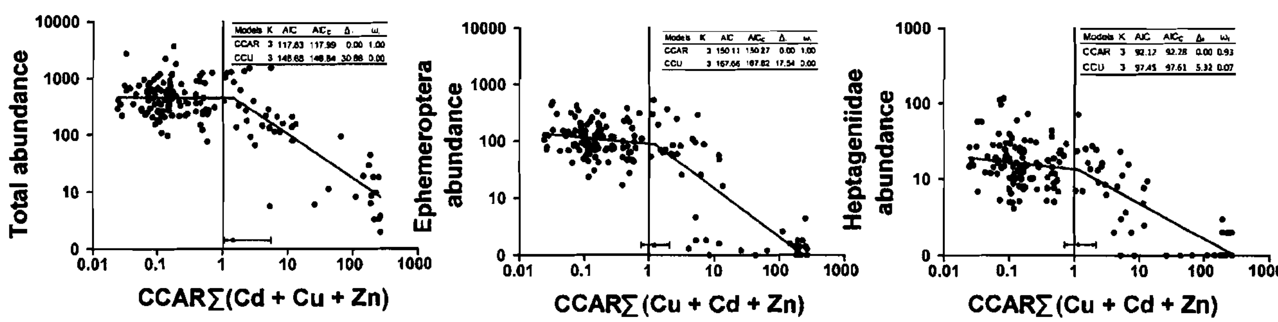

Fig. 4. Piecewise linear regressions showing the relationship between benthic macroinvertebrate community abundance metrics and $\log_{10}$(CCAR) (chronic criterion accumulation ratio). Summaries of model selection analysis are depicted in the table showing Akaike information criteria values. Vertical lines indicate the limit above which adverse effects in aquatic communities are expected. Horizontal segmented lines are the piecewise linear association between richness metrics and CCAR. The 95% confidence interval for piecewise threshold is depicted by horizontal arrows and brackets, and the threshold is identified as the black circle within the confidence intervals. AIC = Akaike information criteria; AIC_c = second-order correction of AIC to account for small ratio of model parameters (K) to observations (n). $\Delta_i = AIC_c$ standardized by subtracting the minimum AIC_c score from each of the candidate models. ω_i = Akaike weight or the probability the model is the best model among the candidate set.

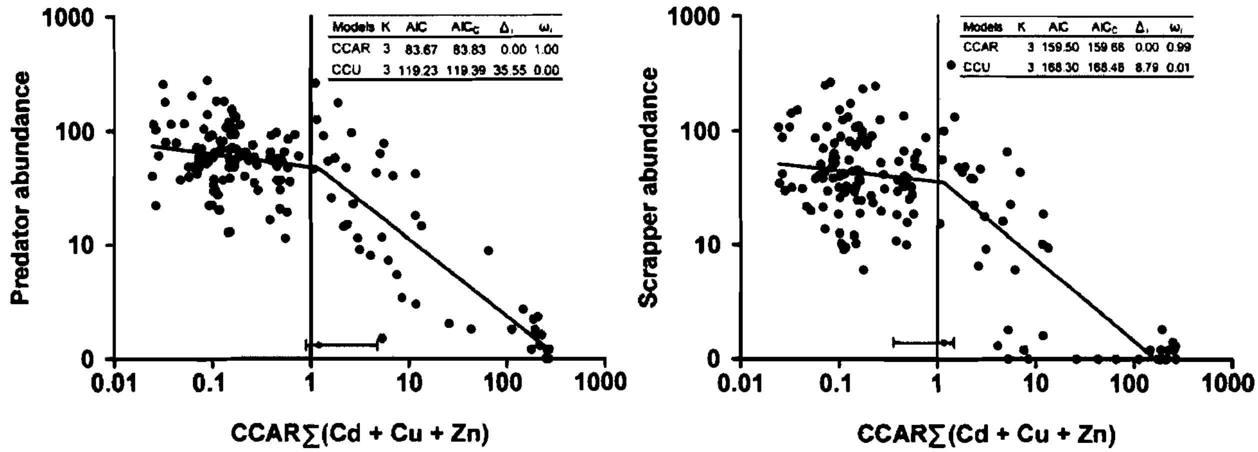

Fig. 5. Piecewise linear regressions showing the relationship between benthic macroinvertebrate community functional metrics and $\log_{10}$(CCAR) (chronic criterion accumulation ratio). Summaries of model selection analysis are depicted in the table showing Akaike information criteria values. Vertical lines indicate the limit above which adverse effects in aquatic communities are expected. Horizontal segmented lines are the piecewise linear association between richness metrics and CCAR. The 95% confidence interval for piecewise threshold is depicted by horizontal arrows and brackets, whereas the threshold is identified as the black circle within the confidence intervals. AIC = Akaike information criteria. AIC_c = second-order correction of AIC to account for small ratio of model parameters (K) to observations (n). $\Delta_1 = AIC_c$ standardized by subtracting the minimum AIC_c score from each of the candidate models. ω_i = Akaike weight or the probability the model is the best model among the candidate set.

Table 5. Results of Mann-Whitney U test showing the mean ± standard deviation metric values between background CCAR[a] and the threshold of chronic toxicity

		Richness metrics		Abundance			Functional	
CCAR[a]	n	Total taxa	EPT[b]	Total taxa	Ephemeroptera	Heptageniidae	Predator	Scrapper
0.1	34	19.7 ± 2.2 A	11.2 ± 1.6 A	591 ± 491	171 ± 114 A	28 ± 28 A	86 ± 62 A	67 ± 60
1	74	18.4 ± 3.3 B	10.5 ± 1.7 B	566 ± 539	114 ± 75 B	16 ± 12 B	64 ± 38 B	54 ± 48
% difference		−7	−6	−4	−33	−43	−26	−19

[a] CCAR = chronic criterion accumulation ratio.
[b] EPT = ephemeroptera + plecoptera + trichoptera.
Capital letters designate significant differences between metric values determined by Mann-Whitney U test ($p \leq 0.05$).

immature insects has been likened to declines in the emergence of adults [43]. Because terrestrial predator density is controlled by, among other things, the productivity of emergent adults [43,44], in-stream disturbances that cause declines in the abundance of benthic macroinvertebrates also limits the food supply to highly dependent terrestrial consumers [43]. Little is known about functional changes to streams in response to disturbance; however, what is known suggests that functional measures are highly sensitive to disturbances, and changes to in-stream function can propagate effects into adjacent terrestrial ecosystems. Future research should focus on the influence of chemical stressors on both in-stream function and also reciprocal changes in the structure and function of dependent terrestrial consumer communities.

CONCLUSIONS

We developed a toxic-unit model using BLM outputs (CCAR) and compared it with CCU to determine which model best predicted benthic macroinvertebrate community responses to trace-metal mixtures. The CCAR was a superior predictor of community responses because it uses the latest knowledge in aqueous geochemistry and physiology of aquatic organisms to predict metal toxicity. Great losses in benthic macroinvertebrate community structure and function were observed near the threshold of chronic toxicity, CCAR at least 1. We report the first measureable losses in benthic community structure and function at a concentration of metals previously thought benign. Our results suggest that CCCs for metals in mixtures are not protective against losses in benthic macroinvertebrate community richness, abundance, and function in Rocky Mountain streams draining mineral deposits.

Acknowledgement—This project was funded by the Central Region Mineral Resources Program (U.S. Geological Survey), and the U.S. EPA STAR program (Grant R829640). Travis S. Schmidt received funding from the USDA National Needs Fellowship Program and the U.S. Geological Survey Mendenhall Post Doctoral Program. Caitlin Balch-Burnett, Sage Betts, Jamie Brennan, Pete Cadmus, Joe Connery, Sigrid Gustafson, Mike Hill, Alicia Louie, Sarah Russell, Dan Scott, David Schnake, David Shaffer, Therese Tepe, Jamie Thayer, Adam Thompson, Monique Adams, Michael W. Anthony, James G. Crock, David L. Fey, Luke McEachron, and Philip L Verplanck all helped to collect and analyze these data. Daren Carlisle, James Lazorchak, Brian S. Cade, Robert E. Zuellig, and David M. Walters provided comments that greatly improved this manuscript. Joseph S. Meyer helped develop the concept of using the BLM as a toxic unit-model. Any use of trade, product, or firm names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

REFERENCES

1. Clements WH, Carlisle DM, Lazorchak JM, Johnson PC. 2000. Heavy metals structure benthic communities in Colorado mountain streams. *Ecol Appl* 10:626–638.
2. Wanty RB, Verplanck PL, San Juan CA, Church SE, Schmidt TS, Fey DL, DeWitt EH, Klein TL. 2008. Geochemistry of surface water in alpine catchments in central Colorado, USA: Resolving host-rock effects at different spatial scales. *Appl Geochem* 24:600–610.

3. Carlisle DM, Clements WH. 1999. Sensitivity and variability of metrics used in biological assessments of running waters. *Environ Toxicol Chem* 18:285–291.
4. Hare L, Tessier A, Borgmann U. 2003. Metal sources for freshwater invertebrates: pertinence for risk assessment. *Human Ecol Risk Assess* 9:779–793.
5. Pagenkopf G. 1983. Gill surface interaction model for trace-metal toxicity to fishes: Role of complexation, pH, and water hardness. *Environ Sci Technol* 17:342–347.
6. Meyer JS. 2002. The utility of the terms "bioavailability" and "bioavailable fraction" for metals. *Mar Environ Res* 53:417–423.
7. U.S. Environmental Protection Agency. 2002. *National recommended water quality criteria: 2002*. EPA 822-R- 02-047. Office of Water, Washington, DC.
8. U.S. Environmental Protection Agency. 2007. Aquatic life ambient freshwater quality criteria—Copper, 2007-Revision. EPA 822-R- 07-001. Office of Water, Washington, DC.
9. Meador JP. 1991. The interaction of pH, dissolved organic carbon, and total copper in the determination of ionic copper and toxicity. *Aquat Toxicol* 19:13–31.
10. Tipping E. 1994. WHAM: A chemical equilibrium model and computer code for waters, sediments, and soils incorporating a discrete site/electrostatic model of ion-binding by humic substances. *Computers and Geosciences* 20:973–1023.
11. Di Toro DM, Allen HE, Bergman HL, Meyer JS, Paquin PR, Santore RC. 2000. Biotic ligand model of the acute toxicity of metals. 1. Technical basis. *EnvironToxicol Chem* 20:2383–2396.
12. Playle RC. 2004. Using multiple metal-gill binding models and the toxic unit concept to help reconcile multiple-metal toxicity results. *Aquat Toxicol* 67:359–370.
13. HydroQual. 2007. Biotic Ligand Model Windows Interface User's Guide and Reference Manual, Ver 2.2.3 HydroQual, Mahwah, NJ, USA.
14. Santore RC, Driscoll CT. 1995. The CHESS model for calculating chemical equilibria in soils and solutions. In Loeppert R, Schwab AP, Goldberg S, eds, *Chemical Equilibrium and Reaction Models*. Soil Science Society of America Special Publication 42. American Society of Agronomy, Madison, WI, pp 357–375.
15. Baron J, McKnight D, Denning AS. 1991. Sources of dissolved and particulate organic material in Loch Vale Watershed, Rocky-Mountain-National-Park, Colorado, USA. *Biogeochemistry* 15:89–110.
16. Prusha BA, Clements WH. 2004. Landscape attributes, dissolved organic carbon, and metal bioaccumulation in aquatic macroinvertebrates (Arkansas River Basin, Colorado). *J North Am Benthol Soc* 23:327–339.
17. Crane M, Kwok KWH, Wells C, Whitehouse P, Lui GCS. 2007. Use of field data to support European Water Framework Directive Quality Standards for dissolved metals. *Environ Sci Technol* 41:5014–5021.
18. Hirst H, Jüttner I, Ormerod SJ. 2002. Comparing the responses of diatoms and macroinvertebrates to metals in upland streams of Wales and Cornwall. *Freshw Biol* 47:1752–1765.
19. Niyogi S, Wood CM. 2004. Biotic ligand model, a flexible tool for developing site-specific water quality guidelines for metals. *Environ Sci Technol* 38:6177–6192.
20. Clements WH, Kiffney PM. 1994. Assessing contaminant effects at higher levels of biological organization. *Environ Toxicol Chem* 13:357–359.
21. U.S. Environmental Protection Agency. 1985. Guidelines for deriving numerical national water quality criteria for the protection of aquatic organisms and their uses. PB85-227049. National Technical Information Service, Springfield, VA.
22. Miller WR. 1999. Geochemical baselines and maps showing acid-neutralizing capacity and potential release of total dissolved solids of stream and spring waters from different rock compositional types from mountainous watersheds in the Gunnison, Uncompahgre, and Grand Mesa National Forest, Colorado. Open-File Report 99-580. U.S. Geological Survey, Denver, CO.
23. Schmidt TS, Church SE, Clements WH, Mitchell K, Fey DL, Wanty RB, Verplanck PL, San Juan CA, DeWitt E, Rockwell B, Klein TL. 2009. Geologic processes influence the effect of mining on aquatic ecosystems. *Proceedings*, 3rd Interagency Conference on Research in the Watersheds, Estes Park, CO, USA, September 8–11, 2008, pp 195–205.
24. Environmental Systems Research Institute. 2006. ArcGIS 9.2, Software. Redlands, CA, USA.
25. Wilde FD, Radtke DB, Gibs J, Iwatsubo RT. 1998. Field measurements. In Wilde FD, Radtke DB, Gibs J, Iwatsubo RT, eds, *National Field Manual for the Collection of Water-Quality Data*. U.S. Geological Survey Techniques of Water Resources Investigations Book 9. U.S. Geologic Survey, Reston, VA.
26. Taggart JE. ed., 2002. *Analytical methods for chemical analysis of geologic and other materials. Open-File Report 02-223. U.S.* Geological Survey, Denver, CO.
27. Moulton SR II, Carter JL, Grotheer SA, Cuffney TF, Short TM. 2000. Methods of analysis by the U.S. Geological Survey national water quality laboratory: processing, taxonomy, and quality control of benthic macroinvertebrate samples. Open-File Report 00-212. U.S. Geological Survey, Denver, CO.
28. Merritt RW, Cummings KW. 1996. *An Introduction to the Aquatic Insects of North America*, 3rd ed. Kendall and Hunt, Dubuque, IA, USA.
29. Ward JV, Kondratieff BC, Zuellig R. 2002. *An illustrated Guide to the Mountain Stream Insects of Colorado*, 2nd ed. University of Colorado Press, Niwot, CO, USA.
30. Cuffney TF. 2003. *User's Manual for the National Water-Quality Assessment Program Invertebrate Data Analysis System (IDAS) Software—Version 3*. Open-File Report 03-172. U.S. Geological Survey, Raleigh, NC.
31. Malmqvist B, Hoffsten P. 1999. Influence of drainage from old mine deposits on benthic madroinvertebrate communities in central Swedish streams. *Water Res* 33:2415–2423.
32. Griffith MB, Lazorchak JM, Herlihy AT. 2004. Relationships among exceedences of metals criteria, the results of ambient bioassays, and community metrics in mining-impacted streams. *Environ Toxicol Chem* 23:1786–1795.
33. R Development Core Team. 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing Vienna, Austria.
34. Sonderegger DL, Wang H, Clements WH, Noon BR. 2009. Using SiZer to detect thresholds in ecological data. *Frontiers in Ecology and the Environment* 7:190–195.
35. Burnham KP, Anderson DR. 1998. *Model Selection and Inference: A Practical Information-Theoretic Approach*. Springer-Verlag, New York, NY, USA.
36. Sokal RR, Rohlf FJ. 2003. *Biometry* 3rd ed. W.H. Freeman, New York, NY, USA.
37. Luoma SN, Rainbow PS. 2005. Why is metal bioaccumulation so variable? Biodynamics as a unifying concept. *Environ Sci Technol* 39:1921–1931.
38. Buchwalter DB, Cain DJ, Clements WH, Luoma SN. 2007. Using biodynamic models to reconcile differences between laboratory toxicity tests and field biomonitoring with aquatic insects. *Environ Sci Technol* 41:4821–4828.
39. Clements WH, Kiffney PM. 1995. The influence of elevation on benthic community responses to heavy metals in Rocky Mountain streams. *Can J Fish Aquat Sci* 52:1966–1977.
40. Clark JL, Clements WH. 2006. The use of in situ and stream microcosm experiments to assess population- and community-level responses to metals. *Environ Toxicol Chem* 25:2306–2312.
41. Brinkman SF, Johnston WD. 2008. Acute toxicity of aqueous copper, cadmium and zinc to Mayfly *Rhithrogena hageni*. *Arch Environ Contam Toxicol* 54:466–472.
42. Carlisle DM, Clements WH. 2005. Leaf litter breakdown, microbial respiration and shredder productions in metal-polluted streams. *Freshw Biol* 50:380–390.
43. Fausch FD, Baxter CV, Murakami M. 2010. Multiple stressors in north temperate streams: lessons from linked forest-stream ecosystems in northern Japan. *Freshw Biol* 55 (Suppl 1):120–134.
44. Burdon FJ, Harding JS. 2008. The linkage between riparian predators and aquatic insects across a stream-resource spectrum. *Freshw Biol* 53:330–346.