

# Predictability of Geochemical Buffering and Runoff Acidification in Spatially Heterogeneous Catchments

JAMES W. KIRCHNER

*Department of Geology and Geophysics, University of California, Berkeley*

PETER J. DILLON AND BRUCE D. LAZERTE

*Ontario Ministry of the Environment, Dorset Research Centre, Dorset, Ontario, Canada*

Catchments are more spatially variable than catchment models assume them to be. Does spatial heterogeneity undermine the validity of spatially aggregated catchment geochemical models? Here we test whether catchment spatial heterogeneity invalidates a recently proposed technique that uses the bulk chemistry of runoff to predict how key water quality variables (e.g., titratable alkalinity and concentrations of base cations, hydrogen ions, and inorganic aluminum) will respond to changes in stream water in acid anion concentrations. We show that these predictions are valid for hypothetical heterogeneous catchments, created by mathematically "mixing" measured runoff chemistry data from chemically diverse streams. The chemical information in the combined runoff yields reliable acidification predictions, without any direct information concerning either the runoff chemistries of the individual source streams or the geochemical properties of their catchments. These results show that catchment acid buffering can be predicted directly from runoff chemistry in spatially heterogeneous catchments, even if the individual source regions cannot be identified and their geochemical characteristics are unknown.

## INTRODUCTION

Catchments' hydrological and geochemical properties are spatially heterogeneous. Does this variability have quantitatively important effects on model predictions of catchments' geochemical behavior? In this paper, we present theoretical and empirical investigations showing that although catchments' geochemical properties are spatially heterogeneous, certain aspects of their geochemical behavior can be reliably predicted.

Catchment geochemical and hydrological properties vary on many spatial scales. Landforms and soil types may vary over scales from tens of meters to kilometers; for example, a catchment might be composed of steep glaciated hillsides with thin, poorly developed soils, as well as deep glacial tills or valley-bottom wetlands. At any particular location, soil geochemistry also varies from one horizon to another. Even at a particular location and within a particular soil horizon, pores and macropores affect the characteristics of the soil matrix at small scales. The chemistry of waters in the interstices between soil particles, and at particle surfaces, may even differ from that of the more mobile water flowing in larger pores, producing chemical variability at very fine scales.

Models of catchment processes do not explicitly represent the spatial heterogeneity of catchment hydrological and geochemical properties on all of the scales that it is known to occur. Even if a completely disaggregated model could be built, collecting the required input data by measuring and mapping the fine-scale variability of hydrological and geochemical parameters would seem to be a practical impossibility. All models of catchment processes therefore average over catchment properties at some scale.

Copyright 1993 by the American Geophysical Union.

Paper number 93WR02202.  
0043-1397/93/93WR-02202\$05.00

Some geochemical models treat the catchment as if it were a single homogeneous unit. Other models divide the catchment laterally into regions or vertically into horizons, which again are treated as homogeneous units. However, the distinction between spatially aggregated and disaggregated models is one of degree, not of kind, because even the most disaggregated models average over considerable spatial variability.

These models require appropriately aggregated measurements of catchments' geochemical and hydrological properties. Making these averaged measurements is inherently difficult, for two reasons. First, measurements of average hydrological and geochemical properties at one scale must attempt to smooth over the spatial variability that exists at finer scales. This variability will often be poorly characterized. Only a tiny fraction of the total volume of the catchments' geochemical and hydrological media can be sampled; it is difficult to make this fraction representative of the whole. For example, a granitic catchment may contain a small region of carbonate-rich rock that makes up the bulk of its acid-buffering potential, but that may be overlooked in a sampling program.

Second, some regions within the catchment may affect runoff chemistry more than others. Points through which most of the catchment's runoff flows, or points where runoff has its last encounter with soils and rocks before flowing into streams or lakes, may be particularly important in determining runoff chemistry. The most appropriate measurement of bulk catchment properties may not be a spatially averaged one, but instead one that weights each point by its influence on runoff [e.g., *Billett and Cresser, 1992*]. Estimating such weighted averages is difficult, because the routing of runoff is often poorly understood even after intensive, catchment-specific field studies. One is therefore faced with an awkward choice between spatially averaged estimates, which

may not be appropriate to the problem at hand, and runoff-weighted spatial averages, which are difficult to estimate.

Because of obstacles like these, average catchment properties are often inferred, rather than measured, by calibrating model coefficients to match the catchment's observed hydrological and geochemical behavior. These calibration exercises usually attempt to adjust the model to reconcile time series data describing the volume and chemistry of precipitation (as an input to the catchment, and to the model) with time series data describing catchment runoff flux and chemistry (as an output from both model and catchment).

The rationale for calibrating models to catchment runoff data is straightforward and appealing. Runoff should reflect the integrated effects of the hydrological and geochemical components of the catchment from which it flows. Moreover, runoff should reflect the properties of the catchment's various components in proportion to their influence on runoff chemistry. Runoff chemistry will not "sample" the geochemical properties of regions that may be extensive but geochemically irrelevant; conversely, it will reflect the effects of regions that may be small but disproportionately influential.

Furthermore, runoff reflects how catchment geochemical media behave under field conditions, rather than the laboratory conditions under which geochemical measurements are often made. For example, in the laboratory exchangeable cations are usually extracted in a concentrated salt solution; it is unclear whether all of these cations are actually available to catchment porewaters, which are typically thousands of times more dilute. In short, the behavior of runoff ought to reflect the "effective" geochemical properties of the catchment. Properly done, model calibration should permit one to extract useful catchment-averaged geochemical information from observed catchment behavior.

However, proper model calibration (i.e., calibration yielding coefficient values that accurately represent bulk catchment properties) requires conditions that may not be met in real life. The detailed time series data required to calibrate typical models are unavailable for most catchments [Eshleman *et al.*, 1992]. Furthermore, if two or more model parameters have similar or offsetting effects on model behavior (or if the data do not display behavior that can constrain them), they will not be individually identifiable through model calibration. Model predictions that depend on the values of these "unidentifiable" parameters will be unreliable.

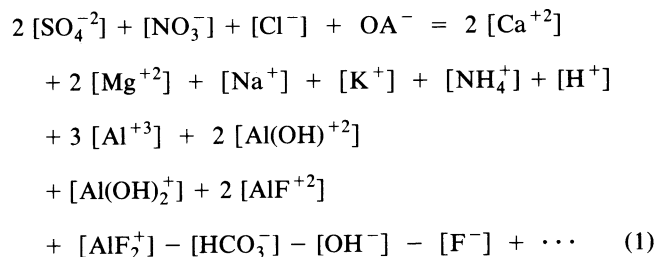
Model parameters are often ill-constrained. Many catchment models are "overparameterized"; i.e., the flexibility introduced by adjustable model parameters far exceeds the power of typical catchment data to constrain them [Hooper *et al.*, 1988; Beck *et al.*, 1990]. Where important model parameters cannot be constrained, model predictions will be highly uncertain [Cosby *et al.*, 1986; Eary *et al.*, 1989; Lee *et al.*, 1990]. Paradoxically, the greater the amount of unconstrained "slack" in model parameters, the better the fit of the model to the calibration data. Because parameter tuning can compensate for, and thus disguise, the effects of unrealistic model structure, successful calibration of highly parameterized models is a weak test of model validity [de Grosbois *et al.*, 1988; Hooper *et al.*, 1990]. Nonetheless, many model users regard the good fit between the calibrated model and the field data as evidence that the model is realistic. Thus overparameterization can inflate model users'

confidence even while it undermines the reliability of model results.

Catchment modelers face a dilemma: a catchment's runoff should reveal its aggregate geochemical and hydrological properties, but typical runoff data are still inadequate to constrain many catchment models. There is a need for techniques that exploit runoff's ability to average over catchments' spatial heterogeneity, without requiring more information than runoff data can provide.

#### A SIMPLE ACID BUFFERING MODEL

Kirchner [1992] has recently proposed a strategy for mechanistically predicting catchment geochemical buffering behavior directly from runoff chemistry, without parameter calibration and its attendant difficulties. This approach draws a direct analytical link between the bulk chemical composition of runoff (which is perhaps the most widely available chemical information for catchments) and catchment response to changes in stream water acid anion concentrations (which one would like to be able to predict for many catchments). In a charge balance for typical acid-sensitive waters,



the left-hand side expresses the combined effects of acid deposition and acid anion retention or release by the catchment. If the sum of acid anions in runoff (the left-hand side of (1)) changes, how will the major ions on the right hand side respond? Catchments that offset changes in acid anions with compensating shifts in base cations will be resistant to acidification. On the other hand, to the extent that the catchment accommodates increases in acid anion concentrations by raising hydrogen ion or inorganic aluminum concentrations, increases in acid loading could make runoff toxic to aquatic organisms.

Kirchner [1992] hypothesized that in catchment runoff, the major ions on the right-hand side are controlled by heterogeneous equilibria (equilibria between the aqueous phase and solid, adsorbed, and gas phases, such as mineral dissolution, ion exchange, and carbonate equilibria). He showed that if these equilibria obey mass action, changes in the equivalent sum of acid anions ( $\Sigma_{\text{acids}}$ ) on the left-hand side of (1) will change each concentration  $C_j$  on the right-hand side by an amount that can be expressed directly in terms of the runoff concentrations themselves:

$$f_j \equiv \frac{dC_j}{d\Sigma_{\text{acids}}} = \frac{z_j C_j}{\sum_{i=1}^n z_i^2 C_i} \equiv \frac{z_j C_j}{h} \quad (2)$$

where  $C_i$  and  $z_i$ ,  $i = 1, \dots, n$ , are the molar concentrations and valences of the  $n$  ions making up the right-hand side of (1),  $C_j$  and  $z_j$  are the molar concentration and valence

of one of the  $n$  ions, whose response to  $\Sigma\text{acids}$  is to be predicted,  $f_j$  is the predicted slope of the buffer curve relating  $C_j$  and  $\Sigma\text{acids}$ , and  $h = \Sigma z_i^2 C_i$  is the heterogeneous buffer term common to all the  $f_j$ .

(We caution the reader that we must necessarily use the term "heterogeneous" in two different senses. Heterogeneous reactions are chemical reactions between two or more different phases, e.g., reactions between ions in solution and those adsorbed to ion exchange media, and heterogeneous equilibria are equilibria of these heterogeneous reactions. However, this paper is also concerned with the spatially heterogeneous geochemistry of catchments, i.e., the spatial variation of geochemical properties and behavior from point to point. The reader should be aware of this semantic ambiguity.)

Equation (2) expresses how much each ion's concentration can be expected to change under shifts in runoff acid anion concentrations or, in other words, the slope of the relationship between  $\Sigma\text{acids}$  and each of the  $C_j$ . Note that heterogeneous equilibrium does not predict that ion ratios in runoff are constant, because the various ions in (2) do not have the same valence. Predicted  $f_j$  values for individual ions can be added, yielding acidification responses for important water quality indicators such as the sum of base cations (SBC), the titratable acid neutralizing capacity (ANC), and the molar sum of inorganic aluminum ( $\text{Al}_i$ ) [Kirchner, 1992; Kirchner et al., 1992]. Equation (2) has been shown to yield quantitatively accurate predictions of how  $\Sigma\text{acids}$  affects calcium, magnesium, inorganic aluminum, hydrogen ions, the sum of base cations, and titratable acid neutralizing capacity in runoff from a variety of catchments in Canada, Norway, and the United States [Kirchner, 1992; Kirchner et al., 1992]. (In addition to this "direct" acidification by changing acid anion concentrations, "progressive" acidification from depletion of exchangeable bases [Kirchner, 1992] can be predicted from extensions to (2), but these extensions will not concern us here). Strictly speaking, this approach predicts catchment response to changing acid anion concentrations in runoff, rather than atmospheric loading to the catchment. However, where empirical relationships or mechanistic models can be used to relate atmospheric deposition and runoff concentrations of acid anions, results from this analysis can be interpreted in terms of catchment response to changes in atmospheric deposition [Kirchner et al., this issue].

Equation (2) contains no equilibrium constants, solid phase activities, or parameters describing the catchments' ion exchange complex, because the equilibrium equations have been inverted to reexpress these coefficients in terms of the solution concentrations themselves (see Kirchner [1992] for details). Like calibration-based modeling approaches, this technique uses geochemical information contained in catchment runoff chemistry, but because the geochemical parameters have been factored out analytically, the parameter identifiability problems outlined above do not arise.

There remains, however, the interesting and potentially important issue of spatial averaging. Both (2) and calibrated catchment models use information contained in runoff that is a mixture from geochemically diverse source regions. Each source region may have a distinct chemical signature and a distinct response to acidification. When flows from each of these regions are mixed, both their chemical signatures and their acidification responses are averaged in some way. How

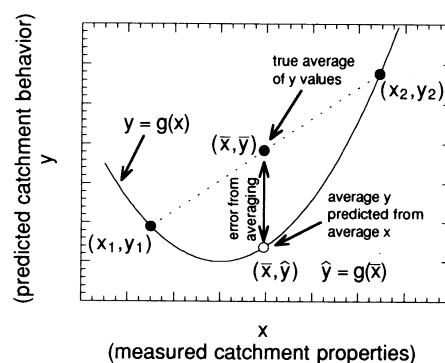


Fig. 1. A simple illustration of the error that results from predicting average catchment behavior ( $y$ ) from averaged catchment properties ( $x$ ) when the governing equation relating catchment properties and catchment behavior  $y = g(x)$  is nonlinear. The individual points  $y_1$  and  $y_2$  yield an average of  $\bar{y}$  (middle solid circle), but when one tries to predict  $\bar{y}$  from  $\bar{x}$ , the average of  $x_1$  and  $x_2$ , one obtains a different value  $\hat{y} = g(\bar{x})$  instead (open circle).

can we tell whether or not predictions derived from this mixed runoff will accurately represent the mixed response to acidification in the spatially heterogeneous catchment from which it flows?

Kirchner [1992] showed analytically that a catchment's spatial variability should have little effect on its response to changes in acid anion concentrations, as predicted by (2) from bulk runoff chemistry. Here, we briefly review the rationale behind this analytical result. We then test this analytical result empirically, by mathematically "mixing" measured runoff chemistry from pairs of intensively monitored streams, creating a series of hypothetical streams that combine flows from geochemically diverse tributaries. We show that even though these hypothetical streams drain spatially heterogeneous "catchments," their acid buffering behavior is geochemically predictable from their bulk chemistry.

#### AVERAGED PARAMETERS IN NONLINEAR MODELS

In both hydrological and geochemical models, one seeks to predict how a catchment will behave (say, how it will respond to changes in climate or acid loading), based on various measured or inferred catchment properties. Typically, catchment behavior (or model output) is a nonlinear function of the catchment properties (or model inputs). A problem arises when averaged properties are used as inputs to these nonlinear relationships, as Figure 1 illustrates. In general, a nonlinear function evaluated at some average input value, shown as the open circle in Figure 1, will not equal the true average of the function's values at each of the individual input values. The mean of the function is not the function of the mean. This problem cannot be solved through better sampling, because it results from nonlinearity in the governing equations rather than uncertainty in the data. In Figure 1, for example, decreasing the uncertainty in the average catchment property  $\bar{x}$  would make the prediction of catchment behavior  $\hat{y} = g(\bar{x})$  more exact, but not more accurate, because the discrepancy between  $\hat{y}$  and the true mean  $\bar{y}$  would remain.

How greatly do these nonlinearities affect predictions of average catchment behavior? In other words, How much

will the predicted average behavior, derived from the averaged properties, deviate from the average of the individual predictions that would have been derived from the spatially disaggregated properties? From Figure 1 one can see that when nonlinear underlying relationships are used to predict average behavior from averaged properties, the magnitude of the resulting distortion depends on two factors: the degree of curvature in the underlying function and the range spanned by the individual data being averaged. Approximating the nonlinear function  $y = g(x)$  in Figure 1 by a second-order Taylor series expansion around the mean  $\bar{x}$ , and averaging this function over various values of  $x$ , one can show that the true mean  $\bar{y}$  is approximately [Ang and Tang, 1975]

$$\bar{y} \approx g(\bar{x}) + \frac{1}{2} \frac{d^2g}{dx^2} \text{Var}(x) \quad (3)$$

Equation (3) coincides exactly with the intuitive understanding outlined above; the discrepancy between  $g(\bar{x})$  and the true mean  $\bar{y}$  is proportional to both the degree of nonlinearity in  $g(x)$ , as defined by its second derivative, and the range of variation in  $x$ , as defined by its variance.

If, as is often the case,  $x$  is not a single variable, but instead an  $m$ -dimensional vector of several variables  $x_1, x_2, \dots, x_m$ , (where the subscript refers to individual variables, rather than individual values of a scalar variable) the simple picture outlined here becomes only slightly more complicated. The multivariate analogue to (3) is simply [Ang and Tang, 1975]

$$\bar{y} \approx g(\bar{x}) + \frac{1}{2} \sum_{j=1}^m \sum_{k=1}^m \frac{\partial^2 g}{\partial x_j \partial x_k} \text{Cov}(x_j, x_k) \quad (4)$$

where  $x_j$  and  $x_k$  are variables that make up the vector variable  $x$ , not individual values of the scalar variable  $x$  used in (3). The only approximation in (3) and (4) is the truncation of the Taylor series at second order; otherwise, the equations are exact and they do not assume that the  $x$  have any particular probability distribution.

Equation (3) or (4) allows one to estimate how much the curvature in the governing equation  $g(x)$  affects estimates  $g(\bar{x})$  of the true mean  $\bar{y}$ . This can be done even if the individual values of  $x$  and  $y$  that make up the means  $\bar{x}$  and  $\bar{y}$  are unknown, so long as one can estimate the approximate variance of  $x$  (or the covariances between its components, if  $x$  is a vector). This is particularly useful in catchment research, where one may have a general idea how much catchment properties vary (or where a worst-case assumption of this variability can be made), but where directly measuring each of the individual  $x$  values can be quite difficult.

#### SPATIAL VARIABILITY UNDER HETEROGENEOUS EQUILIBRIUM

We now apply the formalism described above to estimate how spatial variability in catchment geochemical properties affects predictions of catchment behavior derived from the heterogeneous equilibrium hypothesis. This hypothesis is ideal for exploring the consequences of spatial variability for four reasons. First, the hypothesis itself is both simple and general, and does not assume ad hoc differences in the kinds

of processes controlling runoff chemistry in different parts of the catchment. Second, the only required inputs for (2) are solute concentrations in runoff, which are tangible, measurable quantities. Third, the predicted acidification responses  $f_j$  are simple scalar quantities that can be quantitatively compared with one another; quantifying the differences between two time-dependent simulations (for example, to compare simulations for a spatially uniform catchment and a spatially heterogeneous one) is more difficult. Finally, the dependence of the  $f_j$  on the input data (runoff concentrations) is an explicit mathematical function, not an implicit relationship hidden within various numerical simulation algorithms and calibration procedures. Therefore the derivatives in (4) can be determined explicitly and analytically (in typical simulation models these derivatives could be estimated numerically, but specific numerical estimates would be difficult to generalize).

We apply the formalism above to the following conceptual model of a spatially heterogeneous catchment. We consider a catchment consisting of several source regions with different geochemical characteristics. The source regions may not be individually distinguishable, and their geochemical properties may not be known; we stipulate only that in each source region, heterogeneous equilibrium prevails between the runoff solution and the geochemical media. Water flowing from one region through another may either pass through without chemical alteration, or may equilibrate with the new region. For these purposes, then, the "source region" for a particular fraction of runoff is wherever it picks up its chemical signature. (We also stipulate that the fractional contribution of runoff from each source region remains constant over time; in our companion paper [Kirchner *et al.*, this issue] we explore the implications of relaxing this restriction.)

If heterogeneous equilibrium prevails in each source region, runoff from each source should respond to changes in acid anion concentrations as predicted in (2). However, will the combined runoff of these different source regions also respond as (2) predicts? If the runoff from each source is acidic so that each ion in the charge balance (equation (1)) is either conservative or negligibly small, the concentrations  $C_i$  in the combined runoff should be the average of the  $C_i$  from each source region. If the change in acid anion concentrations ( $d \Sigma \text{acids}$ ) is also similar in each source region, then the acidification response of the  $j$ th ion ( $f_j$ , or  $dC_j/d \Sigma \text{acids}$ ) observed in the combined runoff should be the flow-weighted average of the individual  $f_j$  values. We denote the  $C_i, C_j$ , and  $f_j$  of the combined whole-catchment runoff as  $\bar{C}_i, \bar{C}_j$ , and  $\bar{f}_j$ .

(If the runoff from each source is not acidic, some of the concentrations will not mix conservatively, and thus the concentrations in the mixture will be a nonlinear function of the individual source concentrations. These nonlinearities may compound or offset those embodied in (2), but because the respeciation of nonconservative species cannot be expressed as an explicit function, these effects are omitted here. These effects are, however, included in our empirical tests in the following section.)

Because (2) is nonlinear, the whole catchment average  $\bar{f}_j$  for the combined runoff will not generally equal the geochemical response estimated from the averaged concentrations  $\bar{C}_i$  and  $\bar{C}_j$  in that combined runoff. One can apply (4)

to (2) to estimate the magnitude of the distortion produced by spatial averaging [Kirchner, 1992]:

$$\bar{f}_j \approx \frac{z_j \bar{C}_j}{\bar{h}} \left( 1 + \frac{\text{Var}(h)}{\bar{h}^2} - \frac{\text{Cov}(C_j, h)}{\bar{C}_j \bar{h}} \right) \quad (5)$$

where the second and third terms can be interpreted as fractional adjustments that correct for the effects of spatial variation in  $h$  and  $C_j$ . We expect both of these adjustment terms to be small because, except for very unusual distributions of  $h$  and  $C_j$ , on average  $|h - \bar{h}| < \bar{h}$  and  $|C_j - \bar{C}_j| < \bar{C}_j$ ; thus the variance and covariance terms in (5) are generally much smaller than the products of means by which they are divided. Furthermore, we expect the covariance of  $C_j$  and  $h$  to be positive, because the various runoff concentrations are often positively correlated from point to point within the catchment, and also because  $C_j$  is itself a component of  $h$ . Therefore the two fractional adjustments to  $\hat{f}_j$  in (5) will usually tend to offset one another.

For these reasons, then, we generally expect the whole-catchment acidification response  $\bar{f}_j$  to be closely approximated by  $z_j \bar{C}_j / \bar{h}$ . That is, if heterogeneous equilibrium prevails in each of the source regions, their combined runoff should respond to changes in acid anion concentrations approximately as predicted by (2), using the concentrations measured in the combined runoff itself. In other words, even if the catchment is spatially heterogeneous its buffering behavior should be predictable from its runoff chemistry, as if it were spatially homogeneous. In the following section, we conduct a numerical experiment to test the validity of this theoretical result.

#### AN EXPERIMENT WITH HYPOTHETICAL RUNOFF MIXTURES

Kirchner *et al.* [1992] showed that the heterogeneous equilibrium hypothesis (equation (2)) accurately predicted the response of 18 intensively monitored catchments to large recurrent changes in acid anion concentrations. These acidification episodes (Figure 2) were triggered by acute droughts in the summers of 1987–1990, which permitted oxidation and mobilization of sulfur stored in catchment soils and wetlands [Dillon and LaZerte, 1992]. In many catchments, autumn storms following summer droughts yielded streamflow with very high sulfate concentrations (Figure 2). Acid anion levels remained high for months, falling throughout the year until the next summer's drought. Pronounced changes in base cation, hydrogen ion, and aluminum concentrations, as well as titratable alkalinity, accompanied the acid anion variations (Figure 2). Kirchner *et al.* [1992] tested the heterogeneous equilibrium hypothesis by estimating the expected  $f_j$  from the average runoff concentrations before the first drought (using (2)), then comparing the predicted  $f_j$  with the slope of the relationship between  $C_j$  and  $\Sigma \text{acids}$  observed in the following 4 years.

Here, we test whether the heterogeneous equilibrium hypothesis is similarly successful in predicting the behavior of hypothetical mixtures of runoff from pairs of these geochemically diverse catchments. From the 18 catchments used by Kirchner *et al.* [1992], we selected pairs of streams. We mathematically combined the runoff chemistry time series data from the paired streams (hereafter called "source streams") in equal amounts, forming hypothetical mixtures, hereafter called "merged streams" (Figure 2, Table 1). We

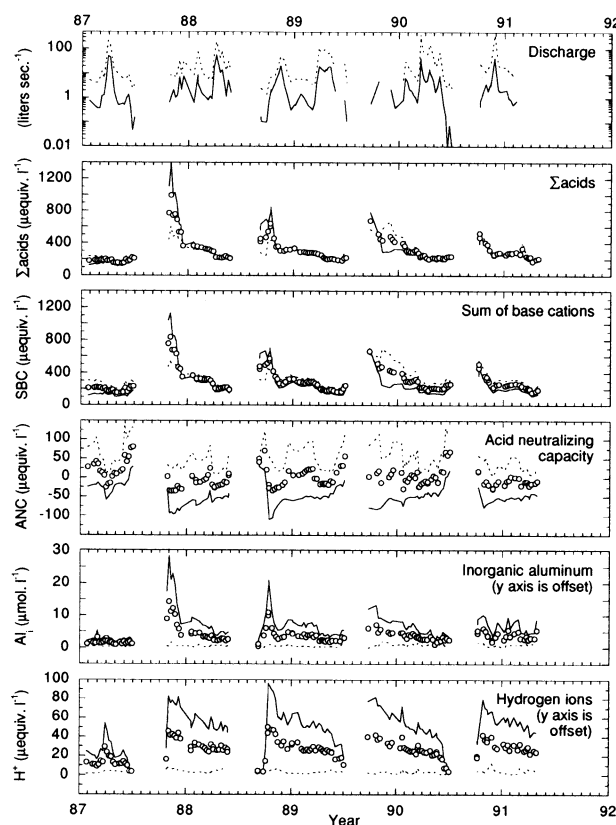


Fig. 2. Effects of summer droughts on chemistry of runoff from catchment 1, Plastic Lake (curves), catchment 5, Harp Lake (dashed curve), and a hypothetical merged catchment (open circles) formed by mathematically mixing runoff from Plastic 1 and Harp 5 in equal proportions (see text). The large data gaps indicate drought periods. Prior to the 1987 drought, runoff chemistry fluctuated within relatively narrow bounds. Droughts in the summers of 1987–1990 were followed by runoff with elevated concentrations of acid anions, base cations, hydrogen ions, and inorganic aluminum, and sharply decreased acid neutralizing capacity. For several months after each drought ended, all of these concentrations remained well outside their predrought ranges. Data are shown only for dates when both streams were sampled.

then tested (2) against the observed behavior of the merged streams, using the same procedure that Kirchner *et al.* [1992] applied to their source streams; we predicted their response to changes in acid anions from their average runoff concentrations before the first drought, then compared these predictions with their observed response over the following four drought years.

The details of our procedure are as follows. For each shared sampling date (each time that both source streams were sampled on the same day or on successive days), we combined the sampled runoff chemistries in a 1:1 ratio. We used a 1:1 mixing ratio to make the combined runoff as different as possible from the two source streams. We averaged the analytical totals of all conservative chemical measurements, including the major base cations, strong acid anions, total inorganic aluminum, total organic aluminum, total fluorine, dissolved inorganic carbon and dissolved organic carbon. We then respecified the weak acids and bases, including hydrogen ions, bicarbonate, aluminum species, and organic anions using the procedures described in Table 1. We varied  $pH$  until the charge imbalance (if any) of

TABLE 1. Mean Preepisode Runoff Concentrations Measured at Study Catchments and Calculated for Hypothetical Runoff Mixtures

	Plastic 4	Harp 6A	1:1 Mixture of Harp 6A and Plastic 4	Range for 18 Source Streams	Range for 18 Merged Streams
pH	4.46	5.24	4.73	4.19–5.54	4.32–5.53
ANC	–41.7	10.5	–15.6	–71.3–54.7	–55.4–36.9
SO <sub>4</sub> <sup>2-</sup>	65.5	129.1	97.3	22.9–146.2	43.6–140.5
NO <sub>3</sub> <sup>-</sup>	0.7	5.6	3.2	0.1–10.1	0.5–8.4
Cl <sup>-</sup>	18.2	10.2	14.2	7.1–30.5	7.6–24.1
Ca <sup>+2</sup>	45.0	113.9	79.4	42.5–122.1	57.9–114.6
Mg <sup>+2</sup>	28.1	52.8	40.4	21.7–64.2	24.8–56.2
Na <sup>+</sup>	24.8	27.8	26.3	16.4–39.8	19.5–33.9
K <sup>+</sup>	10.4	5.9	8.2	1.9–13.1	4.2–9.7
NH <sub>4</sub> <sup>+</sup>	0.3	0.5	0.4	0.2–13.8	0.2–8.5
Al <sub>org</sub> , μmol L <sup>-1</sup>	10.8	3.2	7.0	2.3–10.5	3.1–7.5
Al <sub>i</sub> , μmol L <sup>-1</sup>	6.3	0.6	3.5	0.1–6.7	0.3–4.9
F <sub>T</sub> , μmol L <sup>-1</sup>	3.5	1.9	2.7	2.1–4.0	2.1–3.4
Si, μmol L <sup>-1</sup>	144.7	106.6	125.7	44.7–127.8	57.9–124.5
DIC, mg L <sup>-1</sup>	2.4	2.9	2.7	0.8–5.8	1.8–4.6
DOC, mg L <sup>-1</sup>	16.0	7.3	11.7	2.9–18.9	6.2–18.3
Calculated speciation					
H <sup>+</sup>	34.7	5.7	18.5	2.9–63.9	3.0–48.4
Al <sup>+3</sup>	8.9	0.1	3.2	0.0–9.3	0.0–5.0
Al(OH) <sup>+2</sup>	0.3	0.0	0.2	0.0–1.3	0.0–0.5
Al(OH) <sub>2</sub> <sup>+</sup>	0.0	0.0	0.0	0.0–0.2	0.0–0.1
AlF <sup>+2</sup>	6.1	0.7	4.1	0.1–6.4	0.3–5.4
AlF <sub>2</sub> <sup>+</sup>	0.2	0.2	0.2	0.1–0.6	0.1–0.5
HCO <sub>3</sub> <sup>-</sup>	1.7	12.8	3.7	0.8–36.1	1.2–30.4
OA <sup>-</sup>	71.0	44.3	59.3	6.2–116.6	28.6–104.7

Concentrations (in microequivalents per liter except as noted) are simple means of analytical values for samples collected prior to 1987 drought. Concentrations for hypothetical merged streams were calculated as described in the text. Analytical methods and quality control data are described in detail elsewhere [Ontario Ministry of the Environment, 1983; Locke and Scott, 1986; Janhurst, 1991]. Acid neutralizing capacity (ANC) was measured by Gran titration, and total fluoride (F<sub>T</sub>) was measured by ion-selective electrode with TISAB buffer. Organic and inorganic aluminum (Al<sub>org</sub> and Al<sub>i</sub>, respectively) were measured by catechol violet colorimetry before and after ion exchange [LaZerte et al., 1988]. Aluminum was speciated by the procedure of Schecher and Driscoll [1987]. Organic anions (OA<sup>-</sup>) were calculated from pH and dissolved organic carbon (DOC) by a slight modification [LaZerte and Dillon, 1984] of Oliver et al.'s [1983] method. Bicarbonate was calculated from dissolved inorganic carbon (DIC) and pH. Values reported here are means for matched sample dates and thus differ slightly from values reported by Kirchner et al. [1992], which include all predrought sample dates.

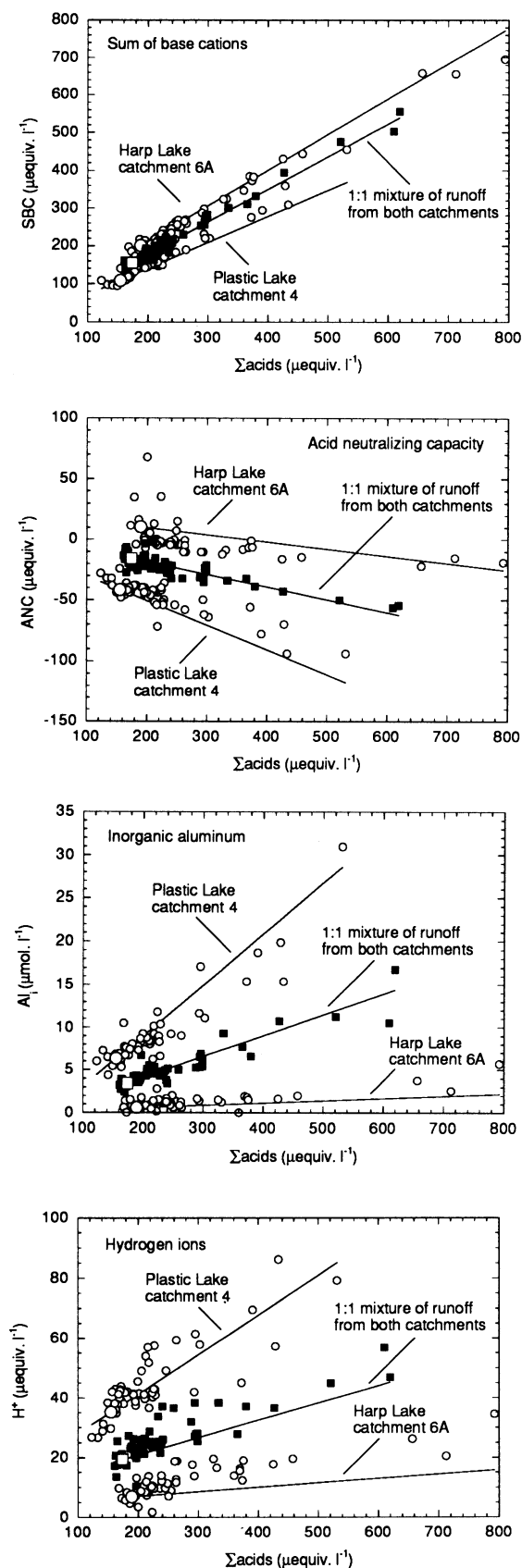
the respiciated solution equalled the average charge imbalance of the two source stream samples (by not enforcing charge balance, we did not force the speciation of weak acids and bases to correct for any charge imbalance due to errors in chemical analyses). Calculated runoff concentrations for nonconservative species often differ from the simple averages of the source stream concentrations (Table 1).

Figure 3 shows the observed response of three streams to changes in acid anion concentrations, compared to the predicted response (lines) estimated from mean predrought runoff chemistry (large open symbols, Table 1) using (2). Two real source streams are shown (open circles), as well as a hypothetical merged stream (solid squares) formed by averaging their runoff chemistry. These two particular source streams were chosen because they are more widely separated in Figure 3 than any other pair of streams would be.

As Figure 3 shows, the heterogeneous equilibrium hypothesis accurately predicts the acidification response of the hypothetical merged stream, even though the two "tributaries" to this stream have utterly different runoff chemistries and acidification responses. It is not surprising that the merged stream responds linearly to changes in Σacids, because it is formed from the two source streams, which

themselves respond roughly linearly to Σacids. Rather, what is significant about Figure 3 is that the slopes of the responses to Σacids in the merged stream are almost exactly those that are predicted, even though the predictions for the merged stream are made solely from its runoff chemistry, without any data describing its geochemically distinct tributaries.

To test the generality of the results shown in Figure 3, we constructed similar plots for 18 hypothetical merged streams formed from 18 pairs selected from the 18 source streams. The 18 stream pairings were selected randomly, subject to several constraints: (1) no pairings were used more than once, (2) each stream was used exactly twice, and (3) each pair had at least 40 shared sampling dates and eight predrought shared sampling dates in common. The range of runoff chemistries encompassed by these 18 merged streams is nearly as wide as that represented by the source streams themselves (Table 1). We repeated the mathematical merging procedure outlined above for each of these randomly selected pairs. We then used (2) to predict how each merged stream, and each source stream, would respond to changes in acid anion concentrations, again using only the average runoff chemistry from before the first drought. We compared these acidification predictions against observed catchment



behavior for each of the 18 actual streams and 18 hypothetical merged streams. We then condensed these results by comparing the predicted slopes (e.g., the slopes of the predicted lines in Figure 3) with the regression slopes of the data for each stream.

Figure 4 shows the predicted slopes and the observed regression slopes of the acidification responses of the real source streams (open symbols) and the hypothetical merged streams (solid symbols). As Figure 4 shows, (2) correctly predicts the overall magnitude of the average acidification response by each ion across the 18 sites, and correctly predicts the differences in response between the different sites. Equation (2) predicts the acidification response of the merged streams approximately as well as it predicts the response of the individual source streams.

In a sense, this result is not at all surprising. It is likely that the individual source streams are themselves derived from spatially heterogeneous catchments. Spatial variability, comparable to that generated artificially in our mixing experiment, is probably present naturally in our source stream data as well. If (2) were vulnerable to the effects of spatial variability, its predictions would probably not be valid for either the real source streams or the hypothetical merged streams. However, instead, the heterogeneous equilibrium hypothesis successfully predicts the behavior of the source streams, despite the spatial heterogeneities that are present. Across our 18 source streams, the predictive reliability of (2) is not diminished as catchment size (and, by implication, complexity) increases, even though catchment area varies by over two orders of magnitude, from 1 to 465 ha (Figure 5).

In the numerical experiments performed above, streams were mathematically merged in a 1:1 mixing ratio. However, in real catchments, changes in hydrologic regime may alter the relative amounts of runoff coming from different source regions, and thus alter the chemistry of catchment runoff. Here, we illustrate how the mixing ratio of two sources can affect the chemical response to changing acid anion concentrations by mathematically merging the two source streams shown in Figure 3 at three different mixing ratios. As Figure 6 shows, the merged runoff's response to acid anion concentrations varies continuously with the mixing ratio. Nonetheless, the observed response to acid anions at each mixing ratio agrees reasonably well with predictions made from (2) using the predrought average runoff chemistry at that mixing ratio (Figures 6 and 7).

Note, however, that the predrought average runoff chemistry at one mixing ratio is not a good predictor of acidification response at another, very different, mixing ratio. That

Fig. 3. (Opposite) Chemical responses of a hypothetical mixed stream (solid squares) and its two source streams (open circles) to changes in acid anion concentrations compared to responses predicted by (2) (lines). The mixed stream has predictable chemical behavior, even though it is formed from two source streams with very different chemical characteristics. The lines shown are predicted relationships, not regression lines. The slopes of the predicted lines were derived from (2), using average runoff chemistry of samples taken before onset of 1987 drought (large open circles). Data for extremely high acid anion concentrations ( $\Sigma_{\text{acids}} > 800 \mu\text{equiv. l}^{-1}$ ) are omitted here and omitted from regression relationships in Figure 4. Only data for shared sample dates (see text) are shown here; the source stream data therefore differ slightly from those presented by Kirchner *et al.* [1992].

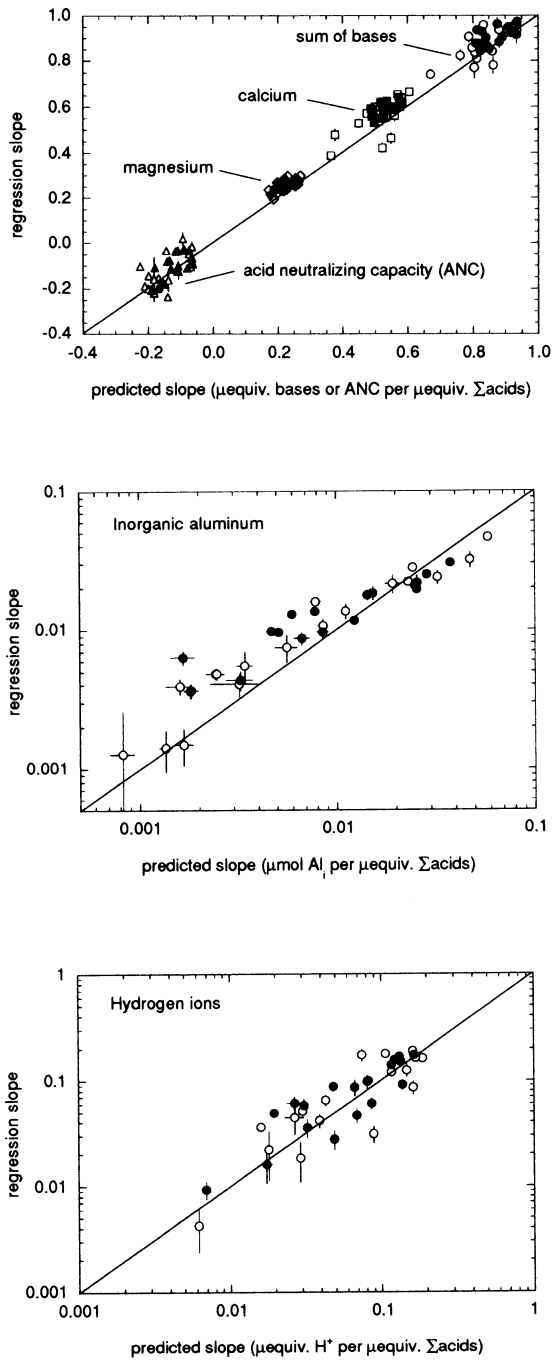


Fig. 4. A comparison of predicted and observed chemical response to acid anion changes in 18 source streams (open symbols) and 18 hypothetical merged streams (solid symbols) for six chemical variables. The diagonal line indicates perfect agreement between predicted and observed buffering behavior. The chemical behavior of the merged streams is as predictable as the chemical behavior of the source streams. The merged streams were formed by mathematically mixing runoff from randomly selected pairs of source streams (see text). For each stream, the predicted line and observed data (as is shown in Figure 3) become a single point here, with the horizontal axis indicating the slope of the predicted line and the vertical axis indicating the slope of the observed data. Predicted slopes, like slopes of lines in Figure 3, were derived from average predrought streamwater concentrations using (2). Observed slopes were calculated by least squares regression, using all samples, both predrought and postdrought. Error bars show standard errors of prediction and standard errors of regression, where larger than plot symbols. Only data for shared sample dates (see text) are shown here; the source stream data therefore differ slightly from those presented by Kirchner et al. [1992].

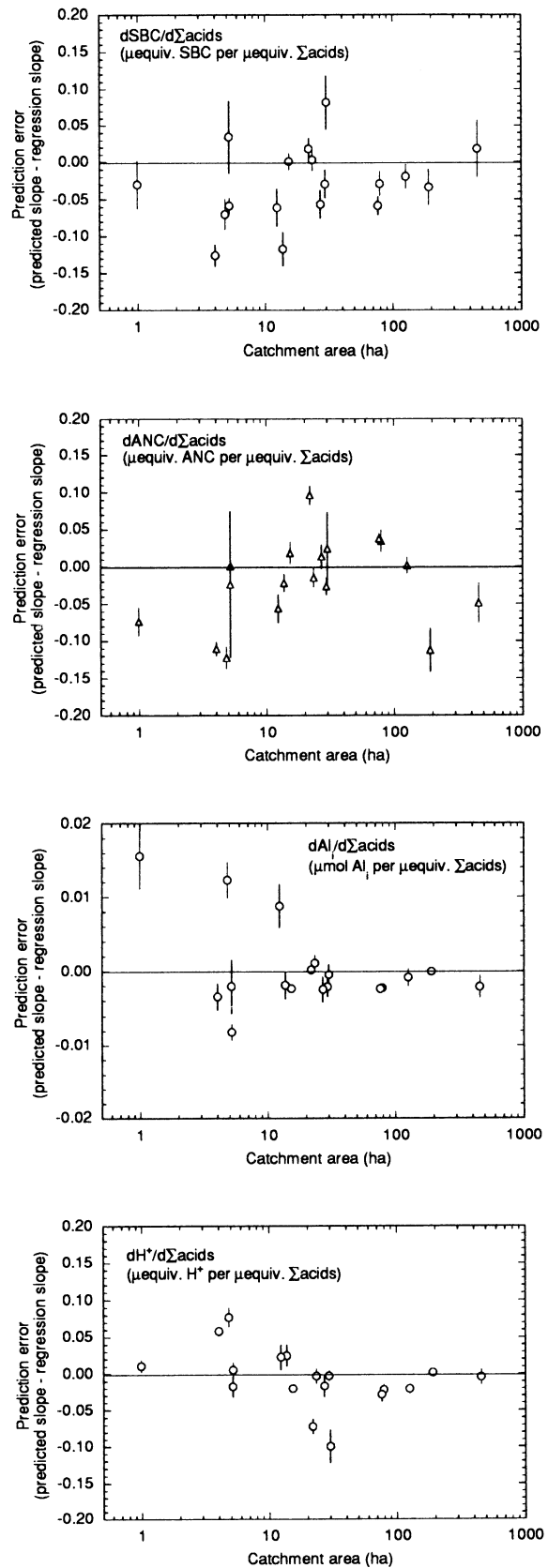


Fig. 5. Prediction error as a function of catchment size for the 18 source streams. Prediction error is defined as the difference between the predicted and observed slopes of response to acid anion changes, as shown in Figure 3. Note that prediction error does not grow with increasing catchment size, which is presumably related to catchment spatial complexity.



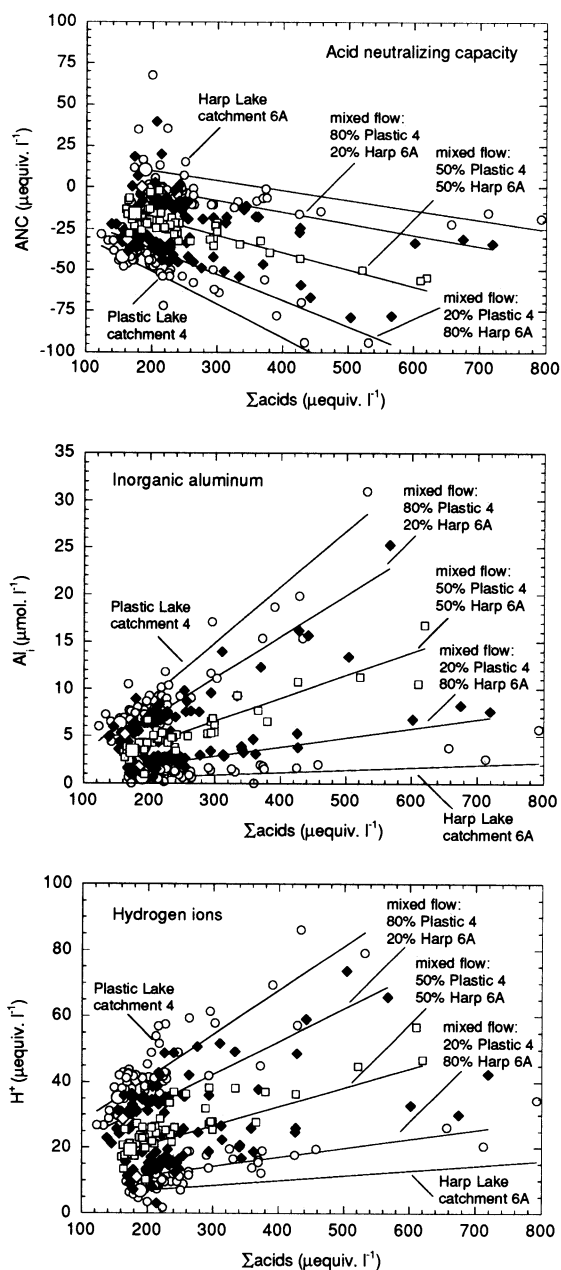


Fig. 6. Predicted and observed chemical response to acid anion changes in two source streams (open circles) and three hypothetical mixtures, formed by mathematically combining the source streams in three different mixing ratios. The predicted lines for each mixing ratio generally describe the observed chemical response for that ratio, but do not describe the trends observed at other mixing ratios. The lines show relationships predicted from average predrought stream water concentrations (large open symbols) using (2). Similar plots are not shown for the sum of base cations (as in Figure 3) because for these data, the different mixing ratios cannot be distinguished visually.

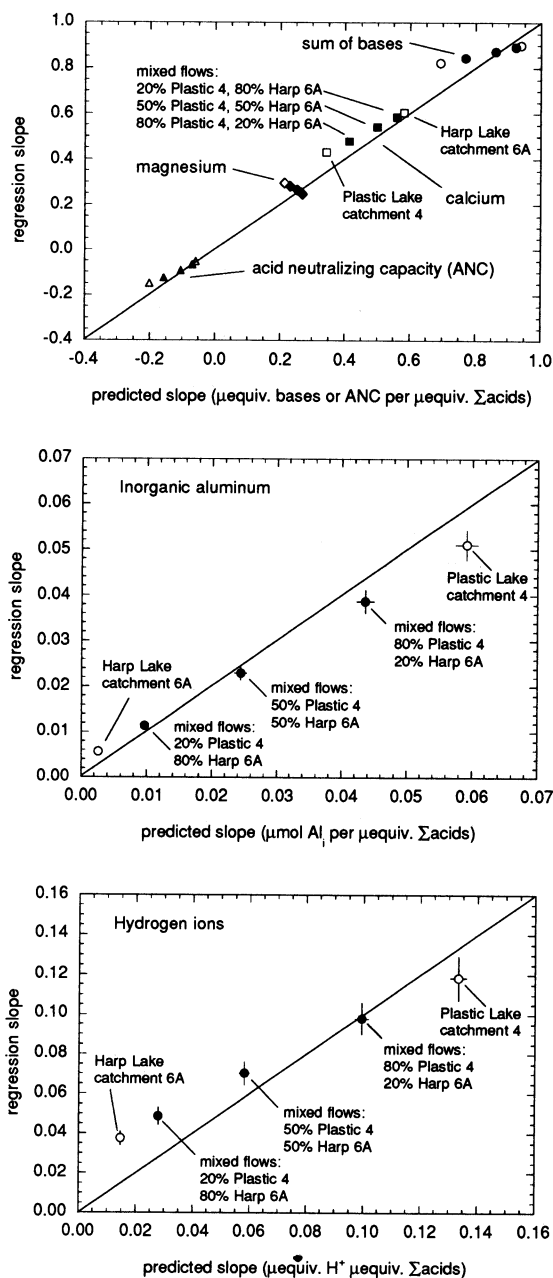


Fig. 7. Predicted and observed slopes of relationships between acid anion concentrations and base cations, acid neutralizing capacity, inorganic aluminum, and hydrogen ions in two source streams (open symbols) and three hypothetical mixtures formed from them at different mixing ratios (solid symbols). The diagonal line indicates that the observed regression slopes coincide with those predicted by (2). Both the predicted and observed slopes for the merged streams form a continuum between the source stream values. For each mixing ratio, the predicted and observed slopes coincide closely, but slopes predicted for one mixing ratio are different from those predicted or observed at other mixing ratios.

is, in Figure 6 the slope of the predicted line for one mixture is not generally a good estimate of the acidification response for the other mixtures. Similarly, in Figure 7 the predicted and observed acidification slopes generally fall close to the 1:1 line, but the predicted slope ( $x$  axis value) for one mixing ratio and observed slope ( $y$  axis value) for another ratio would cross at a point far from the line of perfect agreement. Runoff chemistry observed during one hydrologic regime

(and therefore one mixing ratio) can be a reliable predictor of acidification response under those hydrologic conditions, but it may not be a reliable predictor for other flow regimes.

In some catchments, runoff chemistry responds strongly to changes in discharge (and, by implication, changes in the relative proportions of flow from different source regions). In these catchments, changes in hydrologic regime can obscure the chemical response to changing acid anion concentra-

tions. Imagine, for example, a single data set encompassing a range of discharge corresponding to all of the different mixing ratios shown in Figure 6. If all of these mixing ratios are plotted together, the relationship between acid anions and other chemical variables (for example, ANC,  $Al_i$ , or  $H^+$ ) will be indistinct. Even though  $\Sigma acids$  has clear effects on these variables at each mixing ratio, these effects can be obscured when different mixing ratios are combined. In a companion paper [Kirchner *et al.*, this issue], we describe ways to separate these two effects in runoff chemistry data. Here, we simply note that if catchment runoff chemistry is strongly dependent on discharge, the relationship between acid anion concentrations and other chemical variables may not be obvious in the runoff chemistry data. If heterogeneous equilibrium holds, (2) should correctly describe the chemical response to changing acid anion concentrations, even if a particular data set does not show that relationship clearly. Thus the buffering behavior of spatially heterogeneous catchments is geochemically predictable, even if it is obscured by other factors in catchment time series data.

Our results suggest that catchment spatial heterogeneity does not necessarily invalidate predictions derived from spatially aggregated (or "lumped") geochemical models. This is not to say that such models are necessarily valid; they may suffer from other problems (e.g., unrealistic premises, unidentifiable parameters, or inappropriate formulations for important catchment processes). Therefore our results should not be interpreted as a general endorsement of aggregated geochemical models. Our results suggest only that the assumption of spatial uniformity, while clearly unrealistic, does not by itself compromise geochemical predictions of one particular type: predictions of chemical response to changes in acid anion concentrations under a given hydrologic regime (and thus a given mixture of flows from each source region).

Our results indicate that the chemical response of spatially heterogeneous catchments to changes in acid anion concentrations is functionally equivalent to the chemical response expected for a spatially homogeneous catchment with equivalent bulk runoff chemistry. Equation (2) does not require any direct information about either the runoff chemistries of the individual source streams, or the geochemical properties of the catchments that they drain; it uses only the geochemical information contained in the average bulk chemistry of the merged stream. In other words, when applied to the merged stream, (2) does not "know," nor does it need to know that its input data come from a heterogeneous catchment (nor, indeed, does it need to know anything in particular about the nature of the catchment's heterogeneities). Some aspects of catchments' chemical behavior can be accurately predicted without identifying individual source regions, characterizing their geochemical properties, and mapping the hydrologic flow paths that connect them. This greatly simplifies the problem of predicting the geochemical behavior of spatially heterogeneous catchments.

#### CONCLUSIONS

Because catchment geochemical processes have nonlinear governing equations, spatial heterogeneities may not simply average out in model predictions of catchment geochemical behavior (Figure 1). The magnitude of this effect can be calculated (equations (3) and (4)) from estimates of the

nonlinearity in the governing equations and estimates of the variability in catchment geochemical properties (or measurements used to infer those properties). We use geochemical methods [Kirchner, 1992] that predict catchment acid buffering directly from catchment runoff chemistry to examine the effects of spatial heterogeneity on predictions of catchment geochemical behavior. The distortions introduced by spatial heterogeneity should be relatively small (equation (5)), because the governing equation (equation (2)) is not strongly nonlinear. This theoretical result is tested by mathematically mixing runoff chemistry data from chemically diverse streams, and testing whether the acid buffering behavior of this mixed runoff can be accurately predicted from its bulk chemical composition (Figures 2 and 3).

Acid buffering in mathematically mixed runoff from 18 randomly selected pairs of chemically diverse streams is comparable to a priori geochemical predictions derived from the mixed runoff (Figures 3 and 4). In other words, spatially heterogeneous catchments and spatially homogeneous catchments with equivalent runoff chemistry have functionally equivalent responses to changes in acid anion concentrations. Equation (2) is an equally reliable predictor of catchment acid buffering in the artificially mixed runoff, and in the individual source streams themselves (Figure 4). Furthermore, in the 18 actual source streams, the predictive reliability of (2) is uncorrelated with catchment size and, by implication, catchment spatial complexity (Figure 5). These results show that catchment spatial heterogeneity need not, by itself, undermine the reliability of geochemical predictions of catchment buffering behavior.

In some catchments, changes in hydrologic regime will alter the relative contributions to runoff flowing from geochemically distinct sources, and thus affect runoff chemistry. Where runoff chemistry is strongly dependent on flow, bulk runoff chemistry under a given flow regime (and thus mix of sources) should generally be a reliable predictor of catchment acid buffering under that flow regime, but will not generally be an accurate predictor for other hydrologic conditions (Figures 6 and 7). If catchment runoff chemistry data encompass a wide range of flow conditions, changes in discharge may obscure the catchment's chemical response to changing acid anion concentrations.

The chemistry of each hypothetical "merged" stream contains no information describing the chemical behavior of its individual source streams or the geochemical properties of their catchments. It nonetheless yields valid predictions of whole-catchment response to changes in acid anion concentrations. Predicting the geochemical behavior of spatially heterogeneous catchments does not necessarily require spatially disaggregated geochemical data. Acid buffering can be predicted directly from runoff chemistry in spatially heterogeneous catchments, even if the individual source regions cannot be identified and their geochemical characteristics are unknown.

*Acknowledgments.* Our collaboration was sponsored by a NATO grant to one of us (J.W.K.), and sample collection and analysis were conducted as part of the Acid Precipitation in Ontario Study. We thank R. Reid, R. Girard, and L. Scott for sample collection, A. Neary for supervision of the chemical analyses, M. Futter for database management, and B. A. Roy, N. Christophersen, and three anonymous reviewers for their comments.

## REFERENCES

- Ang, A. H.-S., and W. H. Tang, *Probability Concepts in Engineering Planning and Design*, 409 pp., John Wiley, New York, 1975.
- Beck, M. B., F. M. Kleissen, and H. S. Wheeler, Identifying flow paths in models of surface water acidification, *Rev. Geophys.*, **28**, 207–230, 1990.
- Billett, M. F., and M. S. Cresser, Predicting stream-water quality using catchment and soil chemical characteristics, *Environ. Pollut.*, **77**, 263–268, 1992.
- Cosby, B. J., G. M. Hornberger, E. B. Rastetter, J. N. Galloway, and R. F. Wright, Estimating catchment water quality response to acid deposition using mathematical models of soil ion exchange processes, *Geoderma*, **38**, 77–95, 1986.
- de Grosbois, E., R. P. Hooper, and N. Christophersen, A multisignal automatic calibration methodology for hydrochemical models: A case study of the Birkenes model, *Water Resour. Res.*, **24**, 1299–1307, 1988.
- Dillon, P. J., and B. D. LaZerte, Response of the Plastic Lake catchment, Ontario, to reduced sulphur deposition, *Environ. Pollut.*, **77**, 211–217, 1992.
- Eary, L. E., E. A. Jenne, L. W. Vail, and D. C. Girvin, Numerical models for predicting watershed acidification, *Arch. Environ. Contam. Toxicol.*, **18**, 29–53, 1989.
- Eshleman, K. N., J. P. J. Wigington, T. D. Davies, and M. Tranter, Modelling episodic acidification of surface waters: The state of science, *Environ. Pollut.*, **77**, 287–295, 1992.
- Hooper, R. P., A. Stone, N. Christophersen, E. de Grosbois, and H. M. Seip, Assessing the Birkenes model of stream acidification using a multisignal calibration methodology, *Water Resour. Res.*, **24**, 1308–1316, 1988.
- Hooper, R. P., N. Christophersen, and N. E. Peters, Modelling streamwater chemistry as a mixture of soilwater end-members—An application to the Panola Mountain catchment, Georgia, U.S.A., *J. Hydrol.*, **116**, 321–343, 1990.
- Janhurst, S. (Ed.), 1989 Performance report, Water Quality Sect., Lab. Serv. Branch, Ont. Ministry of the Environ., Rexdale, 1991.
- Kirchner, J. W., Heterogeneous geochemistry of catchment acidification, *Geochim. Cosmochim. Acta*, **56**, 2311–2327, 1992.
- Kirchner, J. W., P. J. Dillon, and B. D. LaZerte, Predicted response of stream chemistry to acid loading tested in Canadian catchments, *Nature*, **358**, 478–482, 1992.
- Kirchner, J. W., P. J. Dillon, and B. D. LaZerte, Separating hydrological and geochemical influences on runoff acidification in spatially heterogeneous catchments, *Water Resour. Res.*, this issue.
- LaZerte, B. D., and P. J. Dillon, Relative importance of anthropogenic versus natural sources of acidity in lakes and streams of central Ontario, *Can. J. Fish. Aquat. Sci.*, **41**, 1664–1677, 1984.
- LaZerte, B. D., C. Chun, D. Evans, and F. Tomassini, Measurement of aqueous aluminum species: Comparison of dialysis and ion-exchange techniques, *Environ. Sci. Technol.*, **22**, 1106–1108, 1988.
- Lee, S. J., K. P. Georgakakos, and J. L. Schnoor, Effects of parameter uncertainty on long-term simulations of lake alkalinity, *Water Resour. Res.*, **26**, 459–467, 1990.
- Locke, B. A., and L. D. Scott, Studies of lakes and watersheds in Muskoka-Halliburton, Ontario: Methodology (1976–1985), *Data Rep. DR 86/4*, Ont. Ministry of the Environ., Dorset Res. Cent., Dorset, 1986.
- Oliver, B. G., E. M. Thurman, and R. L. Malcolm, The contribution of humic substances to the acidity of colored natural waters, *Geochim. Cosmochim. Acta*, **47**, 2031–2035, 1983.
- Ontario Ministry of the Environment, *Handbook of Analytical Methods for Environmental Samples*, Laboratory Services and Applied Research Branch, Rexdale, Ont., 1983.
- Schecher, W. D., and C. T. Driscoll, An evaluation of uncertainty associated with aluminum equilibrium calculations, *Water Resour. Res.*, **23**, 525–534, 1987.
- P. J. Dillon and B. D. LaZerte, Ontario Ministry of the Environment, Dorset Research Centre, Dorset, Ontario, Canada P0A 1E0.
- J. W. Kirchner, Department of Geology and Geophysics, University of California, Berkeley, CA 94720.

(Received November 12, 1992;  
revised August 2, 1993;  
accepted August 4, 1993.)