2002

Application of Classification-Tree Methods to Identify Nitrate Sources in Ground Water

Timothy B. Spruill
United States Geological Survey

William J. Showers
Dep. of Marine Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC

Stephen S. Howe
United States Geological Survey

Ground Water Quality

Application of Classification-Tree Methods to Identify Nitrate Sources in Ground Water

Timothy B. Spruill,* William J. Showers, and Stephen S. Howe

ABSTRACT

A study was conducted to determine if nitrate sources in ground water (fertilizer on crops, fertilizer on golf courses, irrigation spray from hog (*Sus scrofa*) wastes, and leachate from poultry litter and septic systems) could be classified with 80% or greater success. Two statistical classification-tree models were devised from 48 water samples containing nitrate from five source categories. Model 1 was constructed by evaluating 32 variables and selecting four primary predictor variables ($\delta^{15}N$, nitrate to ammonia ratio, sodium to potassium ratio, and zinc) to identify nitrate sources. A $\delta^{15}N$ value of nitrate plus potassium >18.2 indicated animal sources; a value <18.2 indicated inorganic or soil organic N. A nitrate to ammonia ratio >575 indicated inorganic fertilizer on agricultural crops; a ratio <575 indicated nitrate from golf courses. A sodium to potassium ratio >3.2 indicated septic-system wastes; a ratio <3.2 indicated spray or poultry wastes. A value for zinc >2.8 indicated spray wastes from hog lagoons; a value <2.8 indicated poultry wastes. Model 2 was devised by using all variables except $\delta^{15}N$. This model also included four variables (sodium plus potassium, nitrate to ammonia ratio, calcium to magnesium ratio, and sodium to potassium ratio) to distinguish categories. Both models were able to distinguish all five source categories with better than 80% overall success and with 71 to 100% success in individual categories using the learning samples. Seventeen water samples that were not used in model development were tested using Model 2 for three categories, and all were correctly classified. Classification-tree models show great potential in identifying sources of contamination and variables important in the source-identification process.

Nitrate in ground water has been known to be a potential human health problem for more than 50 yr, since Comly (1945) reported that concentrations of nitrate in drinking water could cause methemoglobinemia in infants. A nitrate drinking water standard of 45 mg/L for nitrate (10 mg/L of nitrate, as nitrogen) for United States public water supplies was established in 1962 (United States Department of Health, Education, and Welfare, 1962). This standard has remained in force since 1962 and is the current maximum contaminant level (MCL) for public drinking water supplies (USEPA, 2001).

Some areas of the United States are more likely than others to have high nitrate concentrations in ground water. Susceptibility to nitrate contamination typically is highest in areas with sandy soils (Nolan et al., 1997). Within the Albemarle–Pamlico Drainage Basin of North Carolina and Virginia, the highest nitrate concentrations occurred in areas having sandy soils with relatively low organic carbon content (Spruill et al., 1997; Spruill et al., 1998). Such areas primarily are located in the inner Coastal Plain where dissolved carbon concentrations are less than 3 mg/L. Nitrate concentrations exceeded the 10 mg/L maximum contaminant level in about 5% of the ground water samples from these areas.

To control nitrate contamination in ground water, the nitrate sources must be identified before appropriate and effective management actions can be taken. Ground water can have many nitrate sources, both natural and anthropogenic (Madison and Brunett, 1985; Hallberg and Keeney, 1993; Spalding and Exner, 1993). Rain, forests, grasslands, agricultural lands, organic wastes (e.g., farm manures, sewage sludges, food-processing wastes, and crop residues), row crops, vegetable crops, and livestock production are all potential nitrate sources in ground water.

Nitrogen sources have increased over the last several decades (Smil, 1997; Vitousek et al., 1997). Nationally, nitrogen applications to agricultural lands have increased 20-fold over the last 50 yr, and the most dramatic increases have occurred over the last 30 yr (Puckett et al., 1999). On an annual basis, fertilizer is the largest input of nitrogen to most agricultural systems (Hallberg and Keeney, 1993). In North Carolina, confined feeding operations, particularly with respect to hog production, have increased from 2.2 million hogs in 1990 to more than 10 million hogs in 1999, primarily in the Coastal Plain, making North Carolina the second largest producer of hogs in the United States (Mallin, 2000). In addition, human populations have increased as much as 40% since 1990 in some counties included in this study (United States Census Bureau, 2001). Because of increased nitrogen sources, the many potential regional or local nitrate sources to ground water, and increasing numbers of people in close proximity to these sources, identifying the predominant nitrate sources in ground water may not be easy. Reliable methods are needed that can be used by natural resources scientists and managers to identify sources of nitrate-contaminated ground water.

PREVIOUS STUDIES

Several studies have been conducted over the last 30 yr to identify nitrate sources in ground water (Kreitler,

T.B. Spruill and S.S. Howe, United States Geological Survey, 3916 Sunset Ridge Rd., Raleigh, NC 27607. William J. Showers, Dep. of Marine Earth and Atmospheric Sciences, North Carolina State University, Raleigh, NC 27695-8208. Received 17 Aug. 2001. *Corresponding author (tspruill@usgs.gov).

Published in J. Environ. Qual. 31:1538–1549 (2002).

Abbreviations: CART, classification and regression tree; USGS, United States Geological Survey.

1975; Kreitler and Jones, 1975; Gormly and Spalding, 1979; Fogg et al., 1998) and surface water (Showers et al., 1990). Gormly and Spalding (1979) used isotopes of nitrogen and found that the primary nitrate sources in ground water in Nebraska and corresponding $\delta^{15}N$ range of values were +5 to +9‰ (per mil) for soil nitrogen, −2 to +7‰ for commercial fertilizer, and +10 to +23‰ for livestock. Komor and Anderson (1993) used $\delta^{15}N$ to distinguish nitrate sources in ground water beneath five land-use settings in Minnesota and found that water from wells in livestock feedlots had an average $\delta^{15}N$ concentration of 21.3‰; in cultivated irrigated fields, 7.4‰; in residential areas with septic systems, 6‰; in nonirrigated cropland, 3.4‰; and in natural undeveloped areas, 3.1‰. Several isotope chemists reported that $\delta^{15}N$ concentrations of 10‰ or greater (Kreitler, 1975; Gormly and Spalding, 1979; Aravena et al., 1993; Fogg et al., 1998; Kendall and McDonnell, 1998) indicate that nitrogen from animals is present. In general, $\delta^{15}N$ has been demonstrated to be an effective discriminator between plant or commercial fertilizer–derived nitrate and animal-derived nitrate, but divisions between multiple animal sources and humans are less well defined (Fogg et al., 1998; Kendall and McDonnell, 1998). However, Fogg et al. (1998) indicated that separations between septic and dairy or feedlot sources were possible and, based on their data, septic wastes had a $\delta^{15}N$ signature range from 7.3 to 10.3‰, whereas the $\delta^{15}N$ signature range of the animal sites was from 10 to 14‰.

Thus, although $\delta^{15}N$ of nitrate can be used to distinguish between animal and organic N or inorganic fertilizer-derived nitrate, it has not been successfully used alone to distinguish between subcategories of animal-derived nitrate in ground water. Even coupling $\delta^{15}N$ with other isotopes, such as $\delta^{18}O$, has not been particularly successful for determining differences between animal sources. Nitrate $\delta^{15}N$ data in combination with other water quality variables, such as ions or ionic ratios, however, may be effective in distinguishing animal sources. For example, halogen ratios have been used to identify specific oil-field brines or salt contamination of freshwater aquifers (Whittemore and Pollock, 1979) or to discriminate among precipitation, natural ground water, domestic wastes, and saltwater contamination from evaporites (Davis et al., 1998). By including more variables in the source-identification process, the probability should be greater for successful discrimination among animal sources. Karr et al. (2001) recently coupled the information from both major ion and stable isotope chemistry of ground and surface water to identify sources of nitrate contamination.

MULTIVARIATE STATISTICAL METHODS

Multivariate techniques, both computational and graphical, have been applied to determine the natural phenomena that control ground water quality. Waters associated with specific sources, such as aquifers or petroleum reservoirs, often can be distinguished by using trilinear and pattern diagrams, such as those devised by Piper (1944) and Stiff (1951). Hem (1985) presents several examples of the use of Piper diagrams for distinguishing water composition derived from specific aquifers. These techniques work, in general, because the specific minerals used for source identification either are dissolved by water moving through the rock matrix that composes the natural reservoir or contain connate waters that provide a unique signature of the source. However, for the same reason that makes these diagrams (which use only seven or eight ions) effective at discerning ions derived from a few natural sources, discerning anthropogenic sources with such a limited number of ions becomes considerably more difficult, because of the similarity of concentrations of the same few ions produced by many different natural *and* anthropogenic sources. The use of more sophisticated multivariate techniques, which can incorporate information from many more chemical ions, chemical isotopes, and associated properties to detect unique combinations of variables that identify each source, becomes imperative.

Multivariate statistical methods, capable of distinguishing complex relations among many variables, can be useful for source-identification problems. Alley (1993) presented an excellent overview of multivariate statistical techniques that have been applied to examine phenomena associated with water quality and to understand behavior and spatial patterns of water quality constituents. These techniques include cluster analysis, principal components analysis (PCA), and factor analysis. Steinhorst and Williams (1985) applied multivariate analysis, including analysis of variance, canonical analysis, and discriminant analysis to segregate ground water sources and to differentiate water quality associated with particular aquifers in basalt flows and interbeds in south-central Washington. Multivariate procedures, however, have not been used extensively to determine contamination sources from human activities.

A primary assumption behind this study is that the variability in one or more chemical constituents caused by anthropogenic sources is greater than that caused by other possible natural sources, such as minerals in rocks and soils of the region; therefore, certain constituents can be related to waste-specific sources. The waste-specific sources that often contribute to nitrate contamination are septic-system wastes; fertilizers applied to lawns, row crops, and golf courses; hog wastes leaking from lagoons or sprayed on crops as fertilizer; and chicken wastes applied to crops as fertilizer (Madison and Brunett, 1985; Hallberg and Keeney, 1993).

When the objective of an analysis is to determine into which predefined category a particular observation belongs, discriminant analysis (Davis, 1985) and classification or regression trees (Wilkinson, 2000) are appropriate techniques. Discriminant analysis is a multivariate technique, related to multiple regression, whereby linear equations are found that best discriminate the observations into two or more groups (Wilkinson, 2000). Although either discriminant analysis or classification-tree models are appropriate for the problem of classifying observations into predefined groups, classification-tree techniques have several advantages over discriminant analysis. The primary advantage of classification trees

is that they are graphical and the output is more easily interpreted than strictly numerical methods, such as discriminant analysis (Breiman et al., 1984; StatSoft, 2001). As an example, classification-tree model output is hierarchical (StatSoft, 2001) and produces a visual representation of a dichotomous key, familiar to biologists, that visually and sequentially guides the user through a series of simple if–then statements from the beginning of the tree through a series of subgroups to the final group classification. Other advantages of classification trees over discriminant analysis procedures are that they are nonparametric (Breiman et al., 1984) and can incorporate categorical data, thus making classification-tree methods more versatile with respect to variables that can be included in model development.

After reviewing statistical procedures in available software, classification trees were selected as a versatile tool that can be applied and understood effectively by those who may not have extensive statistical training. Even though many statisticians are not familiar with classification-tree techniques (Wilkinson, 2000), tree models and their development began in the 1960s in the field of social sciences and have, for about the last 20 yr, been extensively used in medicine, marketing, and information management. Regression-tree models (similar to classification-tree models) have only recently been applied to water quality problems. Qian and Anderson (1999) used regression trees to identify factors that affect pesticide concentrations in the Willamette River basin in Oregon. Robertson et al. (2001) used regression trees to identify important environmental variables that affect nutrient concentrations in watersheds in the upper Midwest.

The purpose of this study was to apply tree-based classification methods to (i) determine which water quality variables, both with and without $\delta^{15}N$, could be used to identify the source of nitrate contamination with 80% or better success using selected chemical characteristics of the water sample from five known source categories, and (ii) determine if the chemical characteristics of water samples collected from wells in the North Carolina Coastal Plain and contaminated with nitrate can be used to identify the nitrate source. Ultimately, the intent of this study is to develop and demonstrate the potential of a simple predictive classification procedure that could be used and further developed by environmental scientists and regulators to identify principal nitrate sources present in ground water in a specific geographic area and perhaps apply these procedures to similar environmental problems. Throughout the remainder of this paper, the $\delta^{15}N$ of nitrate will simply be referred to as $\delta^{15}N$.

METHODS

Five common nitrate sources were selected for the analysis—hog wastes sprayed on cultivated fields (Spray), poultry wastes applied as litter (Poultry), septic-system wastes (Septic), inorganic fertilizer applied on golf courses (Golf), and inorganic fertilizer applied on row crops (Crop). Permission was obtained to sample ground water from 4 to 15 locations per category in the Coastal Plain of North Carolina (Fig. 1). Ground water samples were collected directly beneath each source area or, in the case of septic wastes, in the septic field or beneath fields sprayed with septic wastes. Forty-eight ground water samples from 48 wells were included for development of the model.

Wells included in the study were screened to intercept at least the upper 1.5 m of the saturated zone near the water table and were intended to intercept recent (<2 yr old) vertical recharge. The water table of the shallow aquifer usually is located within 3 m of the land surface in the North Carolina Coastal Plain and depth to water ranges between 1 and 3 m below land surface. United States Geological Survey (USGS) wells in the study area intercepted the upper 0.3 to 0.6 m of the saturated zone. In general, areas having sandy soils were selected for sampling to maximize the probability of contamination from nitrate and to ensure that adequate oxygen to maintain nitrate was present. Although only water samples having NO_3–N concentrations greater than 3 mg/L were to be collected (concentration was estimated by using test strips for nitrate), a few samples received from the lab had lower

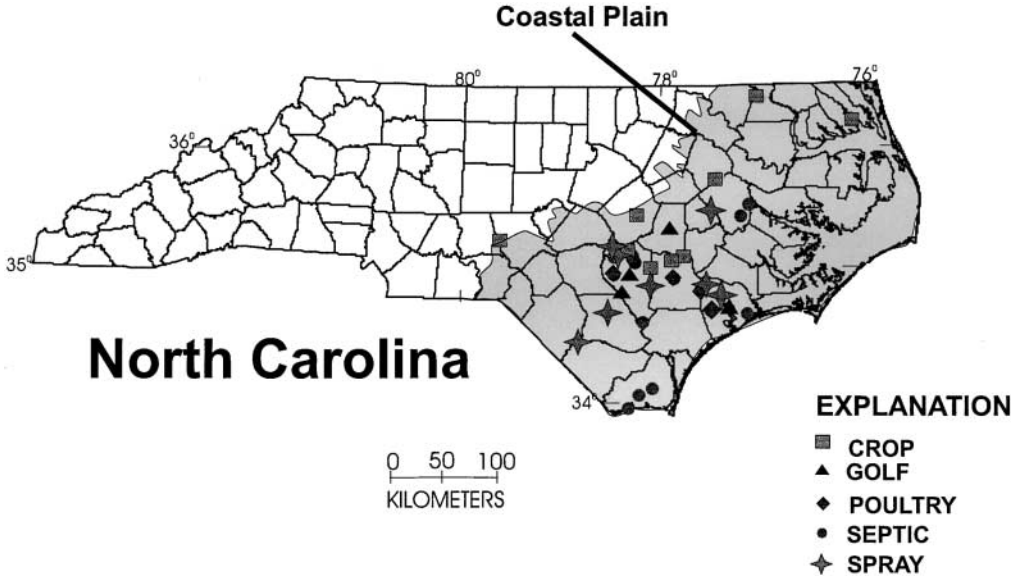

Fig. 1. Locations of sites sampled in North Carolina and nitrate contamination sources.

concentrations. Four samples had concentrations too low (<0.5 mg/L) to analyze $\delta^{15}N$ and were not used. Twenty-six wells were installed and/or used by the North Carolina Department of Environment and Natural Resources (NCDENR) as monitoring wells for a study of pesticides and nitrate in North Carolina ground water (Wade et al., 1997), onsite waste disposal, or other studies. Wells installed by the NCDENR typically were constructed of polyvinylchloride (PVC) with 1.5- to 3-m screens located in the saturated zone of the aquifer beneath the contaminant sources. The USGS installed temporary wells using a minipiezometer assembly (Winter et al., 1988) at 16 of the sites. The minipiezometer was hammered to the desired depth, the 2.5-cm screen extended, and the water sample collected through polytetrafluoroethylene (PTFE) or nylon tubing using a peristaltic pump. North Carolina State University installed six shallow PVC wells that were used in this study.

Each water sample was analyzed for 32 water quality variables that were included in model development (Table 1). Selected water quality data collected from the 48 wells are presented in Table 2. Water samples from 17 additional wells, most with 0.5- to 1.5-m screens, were used to test the resulting models and were collected as part of other USGS and NCDENR–North Carolina Department of Agriculture (NCDA) studies conducted in the study area (Table 3). All water samples collected between August 1996 and February 2000 were filtered through a 0.45-μm capsule filter by using either a peristaltic or submersible pump fitted with either PTFE or nylon tubing. The USGS National Water-Quality Laboratory in Denver, Colorado analyzed major inorganic ions and nutrient species according to methods in Fishman (1993). Either the Stable Isotope Laboratory at North Carolina State University or the USGS Stable Isotope Research Laboratory in Menlo Park, California analyzed samples for $\delta^{15}N$ of nitrate. Determinations of $\delta^{15}N$ were done according to methods presented in Chang et al. (1999) and Silva et al. (2000). Either the USGS National Water-Quality Laboratory or the NCDENR Division of Water Quality Laboratory analyzed the additional 17 well-water samples that were collected as part of the USGS Albemarle–Pamlico Water-Quality Assessment (NAWQA) Program (Spruill et al., 1998) or for the North Carolina Interagency Pesticide Study (Wade et al., 1997).

Table 1. Water quality variables with reporting units included in model development.

Dissolved chemical constituent or property	Unit
Specific conductance	μS/cm at 25°C
pH	units
Ammonia as nitrogen	mg/L
Organic nitrogen plus ammonia	mg/L
Nitrite plus nitrate, as nitrogen	mg/L
Phosphorus	mg/L
Organic carbon	mg/L
Calcium	mg/L
Magnesium	mg/L
Sodium	mg/L
Potassium	mg/L
Chloride	mg/L
Sulfate	mg/L
Silica	mg/L
Fluoride	mg/L
Zinc	μg/L
Bromide	mg/L
Alkalinity	mg/L
$\delta^{15}N$	‰
Calcium to chloride ratio	unitless
Sodium to chloride ratio	unitless
Nitrate to potassium ratio	unitless
Sodium to potassium ratio	unitless
Potassium to chloride ratio	unitless
Calcium to magnesium ratio	unitless
Potassium to phosphorus ratio	unitless
Nitrate to chloride ratio	unitless
Calcium to potassium ratio	unitless
Nitrate to ammonia ratio	unitless
Chloride to sulfate ratio	unitless
Potassium plus $\delta^{15}N$	unitless
Sodium plus potassium	unitless

Two classification-tree models were devised by using the classification and regression tree (CART) procedure (Breiman et al., 1984) on the original 48-sample data set. Model 1 included nitrate $\delta^{15}N$ because it is known to be highly valuable in discriminating animal and fertilizer nitrate. However, $\delta^{15}N$ may not be available because of its cost or because it is not a standard analyte in most ground water monitoring networks. Therefore, all variables, except $\delta^{15}N$, were used in devising Model 2.

The basic idea behind classification-tree models is to create a hierarchical tree of key variables and values based on a sample of objects of known classes (termed the learning sample); the resulting tree is then used to predict classes from another independently obtained sample having the same variables but unknown classes (termed the test sample). Classification-tree procedures employed by many statistical programs begin by separating the initial group composed of all observations (termed the *root node*, which is also a *parent node* or *split node*) into two homogeneous groups (termed *child nodes*) (Fig. 2). The program does this by examining all possible variables and then selecting the best variable (termed the split variable) to split the group into two homogeneous groups (nodes that have the fewest misclassifications or lowest "impurity" and greatest reduction in error from the previous node). The two resulting child groups are now the new parent nodes. The program again splits each of the two new parent nodes into two more child nodes each. This process continues until all of the objects or observations are classified. The groups formed at the end of the tree, which cannot be split any more, form the *terminal nodes* of the tree (Fig. 2).

A variety of tree models including THAID (Morgan and Messenger, 1973), CART (Breiman et al., 1984), FACT (Loh and Vanichsetakul, 1988), and QUEST (Loh and Shih, 1997) are available through several statistical software programs and different tree models may generate different trees according to the classification algorithms employed by the particular model (StatSoft, 2001). Specific splitting algorithms for many of these programs are discussed in Loh and Shih (1997). The CART procedure (Breiman et al., 1984) and a variation, RPART (Therneau and Atkinson, 1997), both used in this analysis, evaluate all variables to determine which variable can make the best split (i.e., the variable that splits the parent group into the two purest child groups) using the GINI index of impurity (i/t) (Breiman et al., 1984). The GINI index is a measure of the total error (also known as deviance, D_i, for classification trees), in any node and is computed by:

$$i/t = 1 - \Sigma_j p^2(j/t)$$

where j is the number of classes in any node t and p is the proportion of the class at the node (Loh and Shih, 1997). Thus, if the first, or root node, contains four classes in equal proportion, then the GINI index is $1 - [(1/4)^2 + (1/4)^2 + (1/4)^2 + (1/4)^2]$ or $1 - 1/4$ or 0.75. A node with only one class (all observations are perfectly classified) would have a GINI impurity index value of $1 - (1)^2$ or 0. The error after the split is the sum of the error of the two resulting child nodes, where D_i (child) = D_i (left child) + D_i (right child). The variable selected would be the one that most reduces the error between the parent and the sum of the error of the two new child nodes:

$$\Delta D_i = D_i \text{ (parent)} - D_i \text{ (child)}$$

Table 2. Selected data used to develop classification-tree models.†

Name	Source	Date	Time	COND	PH	NH4N	KN	NO3N	P	*ZN*	*N15*	*NAK*	*NITK*	*CMR*	*NO3NH4*	*KNO315*	*NAKSUM*
				µS/cm		mg/L				µg/L	‰						mg/L
ED-154	Crop	19960801	1630	399.00	4.90	0.01	0.52	16.00	0.02	NA	2.75	0.72	2.76	11.19	1 142.86	8.55	10.00
ED-146	Crop	19960805	1600	457.00	4.50	0.02	0.38	20.00	0.02	3.50	5.14	0.54	2.50	12.09	909.09	13.14	12.30
PI-586	Crop	19960806	1330	255.00	3.90	0.01	0.20	8.70	0.02	33.00	9.88	7.50	2.42	0.36	621.43	13.48	30.60
PI-587	Septic	19960806	1700	360.00	4.90	0.29	0.51	4.90	0.02	5.10	17.18	52.00	4.90	0.30	16.90	18.18	53.00
DU-150	Crop	19960807	1100	110.00	3.80	0.01	0.20	8.10	0.02	2.60	6.33	2.17	4.50	1.77	623.08	8.13	5.70
DU-151	Crop	19960807	1530	95.00	5.40	0.01	0.20	7.10	0.02	4.50	8.22	6.33	7.89	0.96	710.00	9.12	6.60
DU-147	Crop	19960808	1330	115.00	4.60	0.02	0.20	9.20	0.02	6.30	8.22	2.60	9.20	3.65	575.00	9.22	3.60
DU-148	Crop	19960809	1130	156.00	4.50	0.01	0.20	14.00	0.02	12.00	6.77	0.69	5.38	7.28	1 076.92	9.37	4.40
SA-090	Spray	19960815	1630	282.00	4.20	0.01	0.26	22.00	0.02	7.20	14.57	0.50	1.10	1.50	2 200.00	34.57	30.00
SA-093	Spray	19961119	1015	488.00	4.20	0.02	0.20	52.00	0.01	8.22	23.00	0.54	2.00	2.50	3 466.67	49.00	40.00
SA-092	Spray	19961119	1315	325.00	4.50	0.02	0.20	33.00	0.02	8.22	18.00	0.31	0.92	1.84	2 200.00	54.00	47.00
SA-094	Spray	19961119	1445	38.00	5.40	0.02	0.20	0.19	0.11	8.22	21.00	0.95	0.10	1.56	12.67	22.90	3.70
SA-095	Spray	19961120	1000	220.00	5.40	0.02	0.20	5.00	0.03	8.22	22.00	1.68	1.61	5.94	333.33	25.10	8.30
SA-096	Spray	19961120	1115	281.00	4.70	0.09	0.20	21.00	0.01	8.22	18.50	1.65	4.29	3.84	233.33	23.40	13.00
Co-149	Spray	19980225	1120	310.00	4.10	0.26	0.36	35.91	0.01	NA	12.40	1.40	3.59	2.89	138.12	22.40	24.00
BL-241	Spray	19980519	1530	263.00	5.00	0.94	1.00	21.00	0.02	24.00	12.22	0.58	0.88	0.78	22.34	36.22	38.00
JH-148	Crop	19980728	1230	185.00	4.40	0.01	0.10	15.00	0.04	42.00	2.41	0.92	3.95	4.77	1 500.00	6.21	7.30
GR-085	Spray	19990121	1545	638.00	4.30	0.02	0.39	39.00	0.02	8.10	21.80	3.02	4.53	2.27	1 625.00	30.40	34.60
WA-305	Golf	19990603	1610	163.00	4.80	0.05	0.20	3.10	0.06	8.10	10.60	1.93	1.15	3.25	62.00	13.30	7.90
SA-111	Poultry	19990617	1628	375.00	4.40	0.03	0.25	29.00	0.02	1.90	4.06	0.61	1.04	1.30	966.67	32.06	45.00
SA-112	Septic	19990621	1732	796.00	6.70	1.30	1.70	12.00	0.02	1.00	18.00	17.02	1.28	0.08	9.23	27.40	169.40
SA-113	Poultry	19990623	1900	472.00	4.40	0.40	1.30	13.00	0.02	37.00	17.82	0.42	0.39	3.14	32.50	50.82	47.00
BR-120	Septic	19990701	1120	704.00	7.00	3.70	5.00	0.48	2.10	14.00	14.00	13.00	0.05	0.14	0.13	24.00	140.00
DU-152	Spray	19990708	1217	393.00	4.20	13.00	11.00	27.00	0.02	2.20	17.12	0.15	0.61	1.18	2.08	61.12	50.50
SA-117	Golf	19990715	1305	113.00	4.50	0.02	0.20	4.30	0.02	11.00	5.87	0.91	0.93	1.71	215.00	10.47	8.80
SA-116	Golf	19990715	1640	53.00	4.60	0.02	0.20	0.62	0.02	2.70	8.46	1.77	0.24	0.15	31.00	11.06	7.20
SA-118	Poultry	19990727	1443	454.00	4.00	0.03	0.37	32.00	0.02	1.00	16.33	0.38	1.45	4.74	1 066.67	38.33	30.40
SA-119	Golf	19990730	1505	75.00	4.70	0.01	0.20	2.70	0.02	3.20	2.03	0.95	0.71	1.42	270.00	5.83	7.40
SA-122	Septic	19991022	1210	967.00	4.50	7.00	5.10	74.00	0.03	2.20	9.58	4.73	6.73	1.66	10.57	20.58	63.00
SA-123	Septic	19991022	1605	159.00	4.20	0.02	0.20	13.00	0.02	14.00	5.55	9.67	21.67	1.47	650.00	6.15	6.40
SA-120	Poultry	19991026	1435	958.00	3.90	0.02	0.37	120.00	0.02	2.50	11.97	0.26	1.74	4.11	6 000.00	80.97	87.00
SA-121	Septic	19991028	1200	563.00	3.90	0.06	0.20	41.00	0.02	2.80	11.46	3.29	1.95	0.09	683.33	32.46	90.00
SA-124	Spray	19991116	1515	273.00	5.30	0.02	0.36	4.10	0.02	1.00	20.07	25.00	2.93	0.30	205.00	21.47	36.40
SA-125	Crop	19991117	1300	282.00	4.00	0.03	0.20	22.00	0.02	1.70	4.34	0.54	3.38	6.40	733.33	10.84	10.00
SA-126	Spray	19991118	1630	982.00	5.30	0.02	1.10	43.00	0.10	5.00	36.49	0.39	0.27	0.33	2 150.00	196.49	233.00
ON-013	Spray	19991201	1600	1120.00	6.00	41.00	35.00	24.00	0.18	2.80	28.12	1.81	0.77	1.52	0.59	59.12	87.00
ON-014	Septic	19991206	1430	678.00	7.10	0.02	0.39	8.00	1.10	2.60	18.03	18.18	1.21	0.16	400.00	24.63	126.60
ON-016	Poultry	19991215	1155	1460.00	4.30	0.01	0.86	150.00	0.02	1.00	13.50	1.58	5.77	3.88	15 000.00	39.50	67.00
ON-017	Spray	19991221	1330	1150.00	4.20	0.78	0.89	100.00	0.03	15.00	16.59	1.64	3.03	1.98	128.21	49.59	87.00
DU-005	Poultry	20000119	1630	210.00	4.10	0.02	0.10	27.80	0.01	1.22	10.73	0.89	3.05	3.14	1 390.00	19.23	17.20
DU-004	Poultry	20000120	1400	247.00	5.20	0.02	0.38	12.10	0.00	1.22	26.15	0.72	1.10	2.94	605.00	18.66	18.90
DU-007	Crop	20000215	1430	219.00	4.70	0.02	0.10	13.60	0.01	1.00	7.66	4.50	6.80	1.51	680.00	7.16	11.00
DU-006	Crop	20000216	1045	486.00	5.70	0.02	0.11	27.60	0.02	4.10	5.16	0.77	2.51	6.89	1 380.00	17.09	19.50
SA-002	Crop	20000216	1230	262.00	5.00	0.02	0.55	17.90	0.01	1.90	6.09	3.55	8.95	3.41	895.00	5.18	9.10
SA-003	Crop	20000216	1520	455.00	4.70	0.02	0.11	23.90	0.01	2.40	3.18	0.81	2.88	8.46	1 195.00	14.23	15.00
PK-001	Septic	20000222	1605	3060.00	5.70	0.02	0.62	8.37	0.01	5.60	11.24	14.63	2.04	2.15	418.50	15.34	64.10
HF-003	Crop	20000223	1315	294.00	4.70	0.02	0.16	12.80	0.01	7.00	4.82	1.86	3.66	4.00	640.00	8.32	10.00
MO-038	Crop	20000224	1250	171.00	4.60	0.02	0.10	12.30	0.01	1.60	6.02	0.33	1.37	4.37	615.00	15.02	12.00

† Name, name of well stored in the National Water Information System; Source, principal source of nitrate; COND, specific conductance; PH, pH; NH4–N, ammonia, as nitrogen; KN, ammonia plus organic nitrogen; NO3N, nitrate, as nitrogen; P, phosphorus; ZN, zinc; N15, $\delta^{15}N$; NAK, sodium to potassium ratio; NITK, nitrate to potassium ratio; CMR, calcium to magnesium ratio; NO3NH4, nitrate to ammonia ratio; KNO315, potassium (mg/L) plus $\delta^{15}N$ of nitrate (‰); NAKSUM, sodium (mg/L) plus potassium (mg/L). Constituents selected by model are shown in italic type. All data are available from the United States Geological Survey National Water Information System database, http://water.usgs.gov/nwis (verified 16 May 2002).

Table 3. Data from test sample used to validate Model 2.†

Name	Source	NH4	NO3	CA	MG	NA	K	CMR	NAK	NAKSUM
			mg/L							mg/L
TB-1	Spray	0.01	4.00	12.00	6.20	16.00	5.00	0.39	3.20	21.00
JC-2	Crop	0.02	11.40	7.60	5.50	3.10	4.60	1.77	0.67	7.70
JC-1	Crop	0.02	13.80	11.00	5.80	4.50	4.90	1.29	0.92	9.40
JC-4	Crop	0.01	12.00	7.80	8.50	5.40	3.40	1.57	1.59	8.80
JC-5	Crop	0.05	12.00	7.90	5.20	2.80	3.10	1.86	0.90	5.90
CF-6	Crop	0.01	6.60	15.00	4.40	2.60	7.20	1.69	0.36	9.80
PM-2	Crop	0.01	7.70	26.00	4.70	3.50	6.80	1.34	0.51	10.30
SOW180	Crop	0.09	11.00	64.00	13.00	4.40	1.30	2.95	3.38	5.70
LU-15	Crop	0.16	10.00	18.00	5.50	3.20	4.50	1.72	0.71	7.70
GR-851995	Crop	0.02	10.00	17.00	6.00	8.20	2.30	2.83	0.71	10.50
GR-851999	Spray	0.02	39.00	44.00	15.00	26.00	8.60	2.93	3.02	34.60
L21995	Crop	0.02	4.10	3.50	3.60	2.40	0.70	0.97	3.43	3.10
L22000	Crop	0.02	4.97	2.15	2.47	5.30	0.60	0.87	8.83	5.90
BR-122	Septic	0.03	0.02	0.60	2.40	21.00	1.00	0.25	21.00	22.00
BR-120	Septic	3.70	0.48	16.00	2.00	130.00	10.00	8.00	13.00	140.00
MS4D1	Crop	0.01	2.40	18.00	3.70	5.70	7.40	4.86	0.77	13.10
MS4D2	Spray	0.03	20.00	24.00	7.60	15.00	9.70	3.16	1.55	24.70

† Name, name of well stored in the National Water Information System; Source, principal source of nitrate; NH4, ammonia as nitrogen; NO3, nitrate as nitrogen; CA, calcium; MG, magnesium; NA, sodium; CMR, calcium to magnesium ratio; NAK, sodium to potassium ratio; NAKSUM, sum of sodium and potassium. Data from the United States Geological Survey National Water Information System database, http://water.usgs.gov/nwis (verified 16 May 2002).

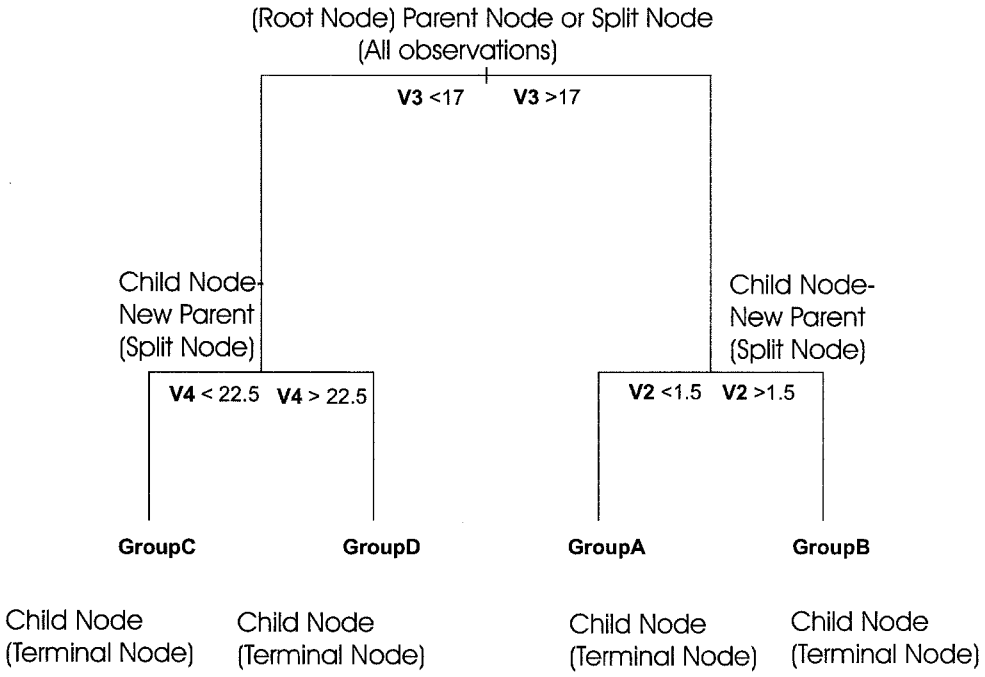

Fig. 2. Diagram of hypothetical classification tree showing node types, split variables, and associated split values.

A succinct description of the GINI index is presented in StatSoft (2001) and Qian and Anderson (1999). It should be noted that the models developed in this paper are not necessarily unique, and it is possible that the model algorithm could select more than one competing variable or split value, particularly with small sample sizes. However, both CART (Breiman et al., 1984) and RPART (Therneau and Atkinson, 1997) were used in the model development process and resulted in very similar models.

An important consideration in devising tree models pertains to the construction of the "right-sized" classification tree (StatSoft, 2001). In essence, how large should the tree be to give the needed predictive accuracy without creating too complex a tree? For example, it may be possible to construct (or "grow") a tree that perfectly classifies all objects, but the resulting tree could be very long and complex, possibly ending with each observation in its own terminal node. A tree that is too short (having too few split nodes) will often have a higher predictive error (or cost) than a more complex tree with more splits and nodes. The issue of when to stop building the tree is a major topic in the classification-tree literature, and good discussions of the principal methods available (including test sample cross-validation, V-fold cross-validation, and global cross-validation) are presented in Breiman et al. (1984) and Statsoft (2001). However, because the intent of this study was largely exploratory in nature and the sample size of 48 observations with five separate groups was very small, a rigorous development of a final fully cross-validated tree model was not the focus of this paper.

In addition to the standard analysis of tree models, the classification success of the terminal nodes of both models (evaluated simply as the percentage of correct classifications of each group) was used to estimate the predictive classification potential of each model, similar to classification matrices produced by discriminant analysis procedures in several commercially available statistical programs. The 48 water analyses shown in Table 2 compose the *learning sample* by which both classification-tree models were constructed. These are the original observations (i.e., water samples with variables selected by the program for construction of Model 1 or Model 2) that form the basis for each model. If the performance were good (80% classification success or better on the learning sample was considered to be acceptable), there would be a basis for adopting the model for practical use or further development to test the model's predictive power and reliability.

Testing on an independent sample and comparing classification success for each category between the learning sample and test sample can be used to demonstrate the practical predictive performance of the model (model validation). However, Model 1 could not be validated by testing with an independent sample, because the primary split variable selected by Model 1 included $\delta^{15}N$, which was not available for analyses of water samples from most wells where the nitrate source was known. All variables identified by Model 2 were available, and the predictive success of Model 2 was validated by using water analyses from an independently obtained test sample of 17 wells not used for Model 2 construction (Table 3) to evaluate model validity. A Kruskal–Wallis test (Conover, 1980) was used when evaluating differences between distributions of model variables among the five sources.

RESULTS AND DISCUSSION

A classification-tree model (Model 1, Fig. 3) was devised by using all 32 variables (including $\delta^{15}N$). The classification tree consists of four splits and five terminal nodes. Only 46 of the original 48 samples were used because of missing zinc data for two of the water samples. The most important variables in this classification tree were potassium plus $\delta^{15}N$ of nitrate (KNO315), nitrate to ammonia ratio (NO3NH4), sodium to potassium ratio (NAK), and zinc (ZN). The resulting classification matrix for evaluating Model 1 performance on the learning sample is shown in Table 4. Source classification of contamination by inorganic fertilizer in both the Crop and Golf categories resulted in 100% correct

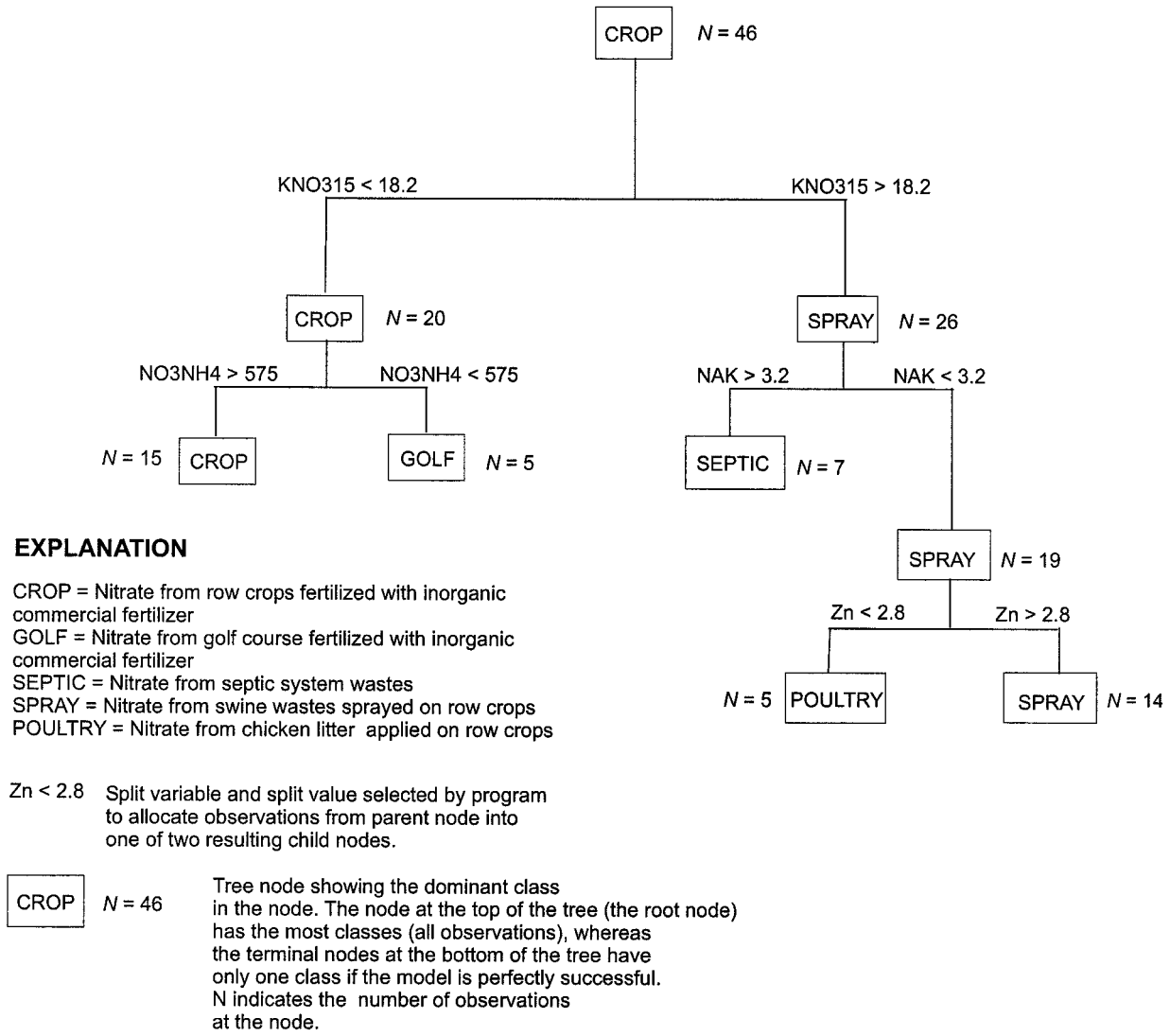

Fig. 3. Classification tree for Model 1 using the predictor variables potassium plus $\delta^{15}N$ of nitrate (KNO315), nitrate to ammonia ratio (NO3NH4), sodium to potassium ratio (NAK), and dissolved zinc (ZN), in micrograms per liter.

placement. The Septic category nitrate sources were classified with 75% success. Water samples from the Poultry category were placed with 71% success. Overall correct classification performance of Model 1 was approximately 88% for all five categories. Because all observations with $\delta^{15}N$ of nitrate were used to develop the model, no independently collected observations (water samples) were available to test model performance.

Model 2 was formulated without $\delta^{15}N$ data. All 48 samples were used in model development. The model that resulted included the sum of sodium plus potassium (NAKSUM), nitrate to ammonia ratio, calcium to magnesium ratio (CMR), and sodium to potassium ratio (Fig. 4). Classification success ranged from 100% for ground water from beneath fertilized golf courses to 71% for water collected from beneath fields fertilized with poultry litter (Table 5a). Overall classification success for the model on the learning sample was about 85%, similar to Model 1. Seventeen samples collected from other areas in the Coastal Plain for three of the five categories were used for validating Model 2. Classification success for Crop, Spray, and Septic categories was 100% (Table 5b).

Application of classification-trees to ground water quality data from eastern North Carolina appears to be very useful in identifying nitrate sources. Model 1 identified four important variables in discriminating between the five groups—potassium plus $\delta^{15}N$ of nitrate, nitrate to ammonia ratio, sodium to potassium ratio, and zinc. Consistent with previous work, much of it

Table 4. Performance of Model 1 on learning sample.

Group and sample size	Group classified by Model 1 (N = 46)					
	Crop	Golf	Spray	Poultry	Septic	Correct
						%
Crop (14)	14					100
Golf (4)		4				100
Spray (13)			12		1	92
Poultry (7)			2	5		71
Septic (8)	1	1			6	75
Overall	15	5	14	5	7	87.6

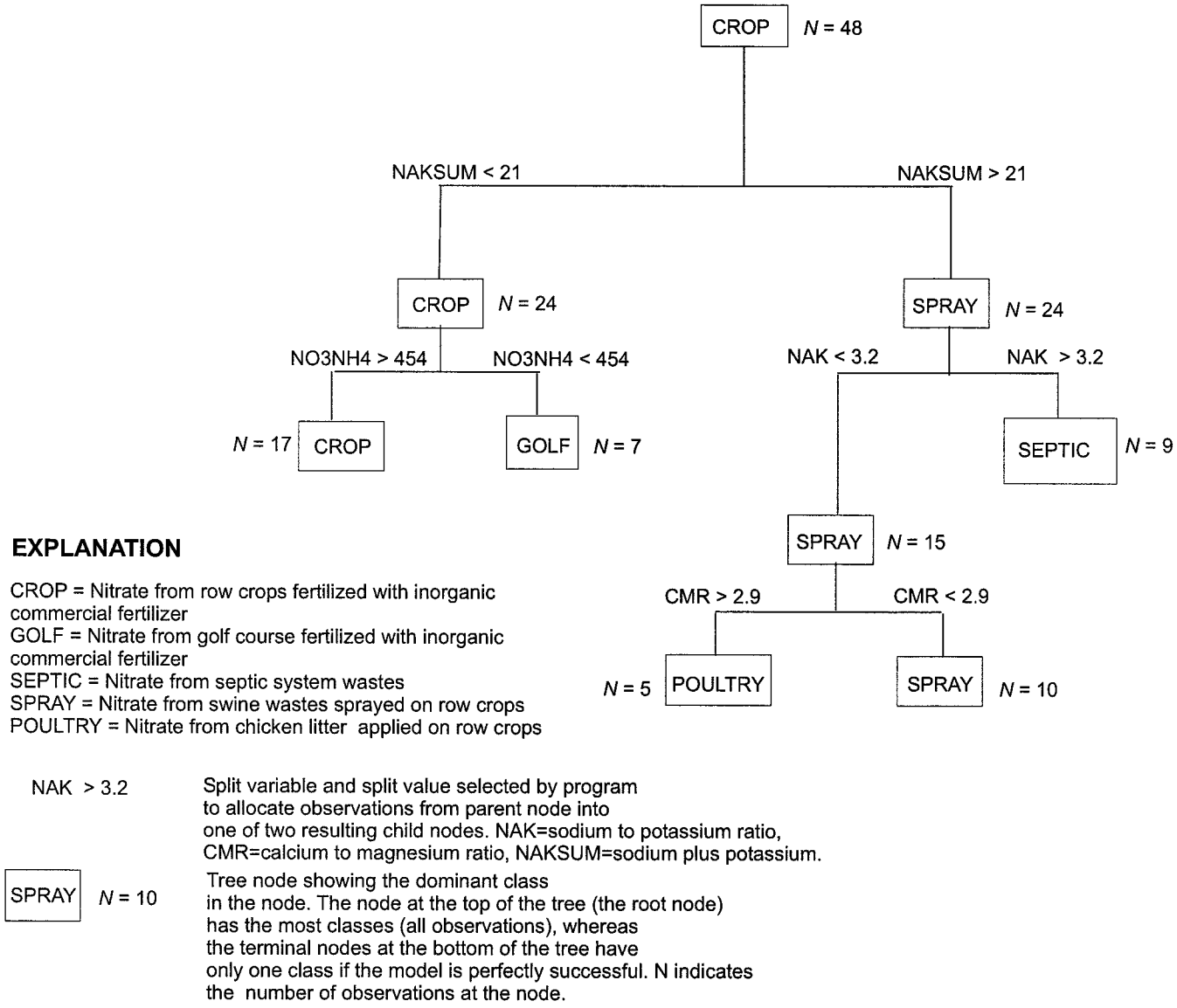

Fig. 4. Classification tree for Model 2 using the predictor variables sum of sodium plus potassium (NAKSUM), nitrate to ammonia ratio (NO3NH4), calcium to magnesium ratio (CMR), and sodium to potassium ratio (NAK).

summarized in Kendall and McDonnell (1998), $\delta^{15}N$ of nitrate is very useful in distinguishing animal sources of N from the other two major environmental sources of N, soil organic N, and fertilizer N. For discussion purposes, another model (not shown) was constructed by using only $\delta^{15}N$, with a resulting model-derived split value (SV) of about 8.5‰ and correctly classified most soil organic and/or inorganic fertilizer sources and animal-based N sources. Based on the learning sample, the model using $\delta^{15}N$ alone was able to correctly classify 17 of 18 fertilizer- or organic N–derived nitrate samples and 29 of 30 animal-source samples. The addition of potassium, in milligrams per liter, to the $\delta^{15}N$ per mil concentrations, however, better separated (i.e., caused less overlap of the distributions) the animal from the inorganic- and/or plant-derived nitrate nitrogen than

Table 5a. Performance of Model 2 on learning sample.

Group and sample size	Group classified by Model 2 (N = 48)					
	Crop	Golf	Spray	Poultry	Septic	Correct
						%
Crop (15)	14				1	93
Golf (4)		4				100
Poultry (7)	2			5		71
Septic (8)	1				7	88
Spray (14)		3	10		1	71
Overall	17	7	10	5	9	85

Table 5b. Performance of Model 2 on test sample.

Group and sample size	Group classified by Model 2 (N = 17)			
	Crop	Spray	Septic	Correct
				%
Crop (12)	12			100.00
Spray (3)		3		100.00
Septic (2)			2	100.00
Overall	12	3	2	100.00

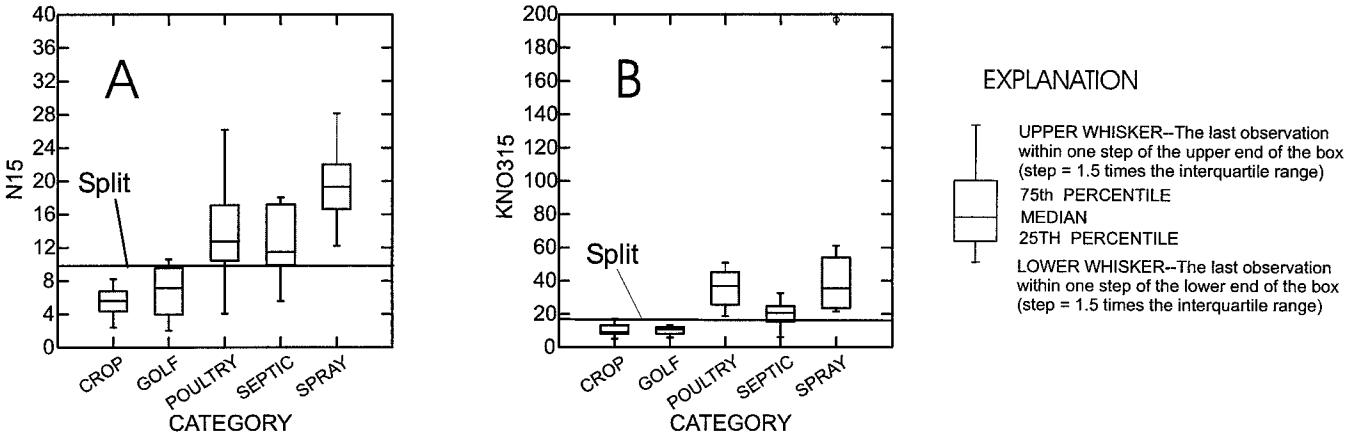

Fig. 5. Distributions of (*A*) N15 (δ^{15}N in per mil) and (*B*) potassium (mg/L) plus δ^{15}N of nitrate (‰) (KNO315) in five source categories demonstrating the effect of adding potassium to increase separation of the animal and fertilizer groups and particularly the Poultry and Septic categories.

δ^{15}N alone, as shown in Fig. 5, and was selected by CART for this data set as the best first split. The primary improvement appears to result from the improved ability to separate poultry from the inorganic- and/or soil organic N–derived nitrogen sources and the Golf category from the animal-derived N sources.

In Model 1, the best discriminator of Golf from Crop samples for the model run shown was the nitrate to ammonia ratio (split value = 575). In general, the Golf water samples had much lower nitrate nitrogen concentrations (median = 2.9 mg/L) than the Crop samples (median = 14.5 mg/L). However, some model runs used nitrate concentrations (model not shown) or other nitrate-related ratios (nitrate to potassium ratio for Model 2) to separate these two groups. The sample size for the Golf category ($N = 4$), however, was so small that it might not be possible to distinguish Crop from Golf categories, unless ground water nitrate concentrations are lower at golf courses compared with those at cultivated fields. Thus, although ground water beneath golf courses appears to have lower nitrate concentrations compared with ground water beneath row crops, many more randomly selected water samples stratified by source would need to be collected to reach such a conclusion.

The best discriminator of septic waste from other animal-derived N sources was the sodium to potassium ratio. Based on information shown in Fig. 6, sodium concentrations in ground water contaminated by septic wastes were higher than those in ground water contaminated by other animal-derived wastes, and the sodium to potassium ratios of septic wastes were significantly higher (median of approximately 14, $p < 0.05$) than other categories investigated (median of all categories < 3). Wilhelm et al. (1994) used sodium concentrations to identify septic-system contamination at a site in Canada. The concentrations were approximately 10 times the background sodium concentration of the ground water (Wilhelm et al., 1994) and the ratio of sodium to potassium in these septic wastes was about 8. Data from Zublena et al. (1993b) indicate that the sodium to potassium ratios for swine lagoon wastes and stockpiled broiler or layer litter (Zublena et al., 1993a) and common fertilizers (Zublena et al., 1991) are all less than 0.5, much lower than the sodium to potassium ratio (approximately 7.5 to 8) indicated by data from Wilhelm et al. (1994) for septic wastes. The sodium to potassium ratio data shown for septic wastes in the North Carolina Coastal Plain in Fig. 6 had a median of about 14 with 75% of the samples exceeding 8, which is comparable

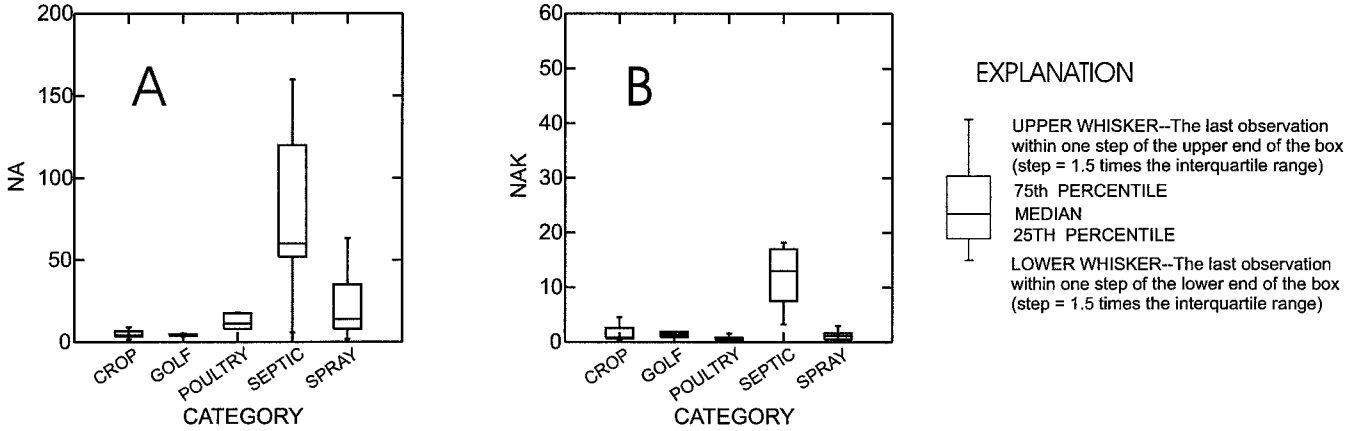

Fig. 6. Distributions of (*A*) NA (sodium, in milligrams per liter) and (*B*) NAK (sodium to potassium ratio, unitless) in five source categories showing increase of separation between septic and the other two animal source categories when NAK is used.

with the ratio shown in Wilhelm et al. (1994). The data from our study suggest that sodium relative to potassium is much higher in septic wastes compared with either of the other animal-derived wastes and may be due to the preponderance of sodium in the typical human diet and the use of salt in water softeners in rural areas. In any case, the sodium to potassium ratio appears to be a good identifier of septic-system wastes within the study area.

After segregating the septic from the poultry and hog-spray wastes (sodium to potassium ratio <3.2, Fig. 3), zinc was useful for further separating the hog and poultry wastes. From the model, a zinc value greater than 2.2 μg per liter (μg/L) indicated hog wastes, whereas values less than 2.2 μg/L indicated poultry wastes. Zinc is added to hog feed as a growth enhancer (National Research Council, 1998) and may be the reason for the higher concentrations observed in ground water samples collected beneath crops fertilized with hog spray.

From the performance data shown in Table 4 for the learning sample, Model 1 appears to be an excellent discriminator of nitrate from inorganic fertilizer on crops, golf courses, and sprayed hog wastes (100, 100, and 92% respectively). Model 1 did not do as well in discriminating between poultry and septic sources, as indicated by the lower classification-success rates (71 and 75% respectively, Table 4). As has been shown by previous researchers, this may be because the $\delta^{15}N$ values of the septic sources have been shown to have a wide range (7.3 to 10‰) that grades into values in both the Crop and Golf categories (Fig. 5), making discrimination difficult. The overlap was not improved by adding potassium (Fig. 5), where the lower tail of the Septic distribution overlaps with the Crop and Golf categories.

Thus, although $\delta^{15}N$ by itself is not particularly successful in separating specific animal sources (Kendall and McDonnell, 1998) and shows no difference between animal categories in the area studied in the Coastal Plain of North Carolina (Fig. 5), using it in combination with other isotopes (such as $\delta^{18}O$, as suggested in Kendall and McDonnell, 1998) or ions, as demonstrated by results shown in this paper, can potentially segregate by animal-source category. An advantage of using major ions, as opposed to various isotopes, is related to the generally lower cost of the analysis for major ions. Although major ions alone can be used effectively in eastern North Carolina and probably most areas where the specific conductance of the shallow ground water is 350 μS/cm or less, specific models probably will need to be devised for areas where specific conductance is typically greater than this. Such areas include coastal areas and parts of the western and midwestern United States where evaporite deposits or saltwater intrusion occurs. In these areas, $\delta^{15}N$ is probably the best indicator of nitrate sources. In such areas, further separation of nitrate sources by using major ions may be difficult or require trace elements or other isotopes.

Nevertheless, in North Carolina and perhaps other areas of the East Coast where shallow ground water has relatively low dissolved solids, major ions can be used effectively to identify sources, as indicated by results shown for Model 2 (Fig. 4, Table 5a). In this model, sodium plus potassium, in mg/L, was found to be an excellent indicator of inorganic and/or soil organic N and animal-derived nitrate sources, with only one crop fertilizer–derived water sample misclassified as septic-derived N and one septic-derived sample classified as nitrate from an inorganic fertilizer source (Table 5a). The overall classification success rate for Model 2 on the learning sample was 85%. The primary distinguishing characteristic of water samples from golf courses was the low nitrate concentration, although statistical limitations of its use for this purpose have been mentioned already. The nitrate to ammonia ratio was used by Model 2 (as in Model 1) to best distinguish the two categories, although the split value (454) was lower in this model. The calcium to magnesium ratio (split value = 2.9) was best used to distinguish poultry from hog spray, and sodium to potassium ratio was best used to distinguish septic from hog spray. The performance of the calcium to magnesium ratio in identifying poultry sources was identical to the performance of zinc in Model 1 (71% success, Table 4). Calcium and magnesium may be easily leached in the North Carolina Coastal Plain, where the cation exchange capacity (CEC) is typically low (<2 $cmol_c/kg$). The mobility of cations may be greatly enhanced in much of the Coastal Plain, which may allow for their use in source identification in this and other areas having low CEC.

Although additional samples would be desirable in formulating a more precise model, both Model 1 and Model 2 appear to be effective in identifying nitrate from specific waste sources, at least for inorganic fertilizer-derived nitrate (Crop, Golf) and animal-derived nitrate (Spray and Septic) categories. Model 2 was tested using 17 water samples that were not used in model formulation, yielding a 100% classification success rate for the three categories (Crop, Septic, and Spray) for which data were available. The reliability of the model is further substantiated in that one well (GR-851995; Table 3) in the test data set sampled in 1995 was identified as an inorganic fertilizer source and in 1999 was identified as a hog-waste spray source (GR-851999; Table 3). Hog spray was indeed used after 1995 for fertilizing crops grown in this field and the model correctly identified nitrate sources for each time period. The water sample from L2 in 1995 (L21995; Table 3) indicated inorganic fertilizer and/or soil organic nitrogen as a source and again in 2000 (L22000; Table 3). This area is not affected by spray and is upgradient from fields that received spray. In addition, two drainage ditches (MS4D1 and MS4D2; Table 3) drain fields fertilized with inorganic fertilizer and hog spray, respectively, and were identified correctly by the model.

A significant finding of this study was that, with the exception of nitrate, no anion was identified as an important classification variable. These results suggest that although anions generally are more mobile in water, they do not differ significantly in concentration among source categories in shallow ground water of the North Carolina Coastal Plain. Even nitrate was found to be

important only in distinguishing the fertilizer from crop and golf courses; of the four golf course samples used, all had lower nitrate, which may or may not be generally representative of golf courses. No significant differences were found among categories for sulfate ($p > 0.05$), and chloride in the Septic category was significantly higher ($p < 0.05$) than the Crop, Golf, and Poultry categories, but not the Spray category ($p > 0.10$), which explains why sodium was selected by the model.

CONCLUSIONS

There are many possible applications of the classification-tree models presented in this paper. Some of these applications include determining nitrate sources in wells that appear unusual (i.e., determining the source of high nitrate concentrations in the vicinity of other wells that have much lower concentrations); determining the principal source of high nitrate where multiple sources may be contributing (septic tank vs. nearby chicken or crop-farming operations); and evaluating effectiveness of management actions (i.e., eliminating a source of contamination, such as a leaking sewer or spray application).

The classification-tree models developed in this study demonstrate that they are useful in identifying variables that are important in the source-identification process and that $\delta^{15}N$, dissolved calcium, magnesium, sodium, potassium, nitrate, ammonia, and zinc are potentially useful in identifying dominant nitrate sources in ground water in sandy recharge areas of the Coastal Plain. Anions in general were not identified in the modeling process as important in discriminating nitrate sources in the study area, although further work and larger sample sizes will be needed to verify this. Specifically, although the classification-tree models may be applied as presented here, they are not unique or the only models possible, and additional ground water samples collected throughout the North Carolina Coastal Plain will be needed to better identify particular nitrate sources and improve the models, particularly for septic and poultry sources. Although this process may lead to more complicated tree models, it could also result in more precise classifications.

Although the simple models presented in this paper may be suitable for shallow aquifers in the North Carolina Coastal Plain and much of the middle Atlantic Coastal Plain, specific applications that may include other sources or contaminants (i.e., gas stations, landfills, etc.) in other areas would require the gathering of data from additional ground water sites with samples to be collected from known sources, such as was done in this study. Classification-tree models are widely available in many statistical computer packages, are relatively easily implemented and interpreted, and appear to classify sources at a level of reliability that can be practically useful.

The nitrate-source identification techniques used here appear to be generally useful in the Coastal Plain of North Carolina and possibly other areas having shallow ground water and low specific conductance, although further research is necessary to address questions about resulting mixtures, influence of oxidation–reduction conditions in the aquifer, degradation or sorption of particular chemical indicators along flow paths, and interference with high background concentrations of ions that are used as indicators. As has been noted already, $\delta^{15}N$ appears to be a reliable indicator under conditions where other chemical indicators would not be as effective. Thus, inclusion of $\delta^{15}N$ in analyses is almost always advantageous for identification of sources and in establishing model plausibility. Data presented in this paper also demonstrate that routine inclusion of major ions as part of water quality studies that are not specifically directed at understanding the geochemistry can yield information that is highly useful, if not necessary, for meaningful data interpretation.

ACKNOWLEDGMENTS

This project is a cooperative effort between the United States Geological Survey (USGS), the North Carolina Department of Environment and Natural Resources (NCDENR), and the United States Environmental Protection Agency (USEPA). Thanks to the many landowners, farmers, golf course managers, and others in eastern North Carolina who allowed access to their property. Special thanks to the USGS National Water-Quality Assessment Program; Song Qian, The Cadmus Group, Durham, NC; Diana Rashash, North Carolina State University Cooperative Extension, Onslow County; Wendell Gilliam, North Carolina State University; and Ray Milosh, Carl Bailey, Elizabeth Morey, Ted Mew, and Paul Dahlen, NCDENR. Finally, thanks to all of the reviewers of this paper who made many helpful comments and suggestions.

REFERENCES

Alley, W.M. 1993. Regional ground-water quality. Van Nostrand Rheinhold, New York.

Aravena, R., M.L. Evans, and J.A. Cherry. 1993. Stable isotopes of oxygen and nitrogen in source identification of nitrate from septic systems. Ground Water 31:180–186.

Breiman, L.J., J.H. Friedman, R.A. Olshen, and C.J. Stone. 1984. Classification and regression trees. Chapman and Hall/CRC, New York.

Chang, C.C., J. Langstron, M. Riggs, M.H. Campbell, S.R. Silva, and C. Kendall. 1999. A method for nitrate collection for $\delta^{15}N$ and $\delta^{18}O$ analysis from waters with low nitrate concentrations. Can. J. Fish. Aquat. Sci. 56:1856–1864.

Comly, H.H. 1945. Cyanosis in infants caused by nitrates in well water. JAMA 129:112–116.

Conover, W.J. 1980. Practical nonparametric statistics. John Wiley & Sons, New York.

Davis, J.C. 1985. Statistics and data analysis in geology. John Wiley & Sons, New York.

Davis, S.N., D.O. Whittemore, and J. Fabryka-Martin. 1998. Uses of chloride/bromide ratios in studies of potable water. Ground Water 36:338–350.

Fishman, M.J. (ed.) 1993. Methods of analysis by the U.S. Geological Survey National Water Quality Laboratory—determination of inorganic and organic constituents in water and fluvial sediments. Open-File Rep. 93-125. United States Geol. Survey, Reston, VA.

Fogg, G.E., D.E. Rolston, D.L. Decker, D.T. Louie, and M.E. Grismer. 1998. Spatial variation in nitrogen isotope values beneath nitrate contamination sources. Ground Water 36:418–426.

Gormly, J.R., and R.F. Spalding. 1979. Sources and concentrations of nitrate-nitrogen in ground water of the Central Platte Region, Nebraska. Ground Water 17:291–301.

Hallberg, G.R., and D.R. Keeney. 1993. Nitrate. p. 297–322. *In* W.M. Alley (ed.) Regional ground-water quality. Van Nostrand Rheinhold, New York.

Hem, J.A. 1985. Study and interpretation of natural water. Water-Supply Paper 2254. United States Geol. Survey, Reston, VA.
Karr, J.D., W.J. Showers, J.W. Gilliam, and A.S. Andres. 2001. Tracing nitrate transport and environmental impact from intensive swine farming using delta nitrogen-15. J. Environ. Qual. 30:1163–1175.
Kendall, C.A., and J.J. McDonnell. 1998. Isotope tracers in catchment hydrology. Elsevier, Amsterdam.
Komor, S.C., and H.W. Anderson, Jr. 1993. Nitrogen isotopes as indicators of nitrate sources in Minnesota sand-plain aquifers. Ground Water 31:260–270.
Kreitler, C.W. 1975. Determining the source of nitrate in ground water by nitrogen isotope studies. Rep. of Investigations 83. Bureau of Econ. Geol., Univ. of Texas, Austin.
Kreitler, C.W., and D.C. Jones. 1975. Natural soil nitrate: The cause of nitrate contamination of ground water in Runnels County, Texas. Ground Water 13:53–61.
Loh, W.-Y., and Y.-S. Shih. 1997. Split selection methods for classification trees. Stat. Sinica 7:825–840.
Loh, W.-Y., and N. Vanichsetakul. 1988. Tree-structured classification via generalized discriminant analysis (with discussion). J. Am. Stat. Assoc. 83:715–728.
Madison, R.J., and J.O. Brunett. 1985. Overview of the occurrence of nitrate in ground water of the United States. *In* National Water Summary 1984—Hydrologic events, selected water-quality trends, and ground-water resources. Water-Supply Paper 2275. United States Geol. Survey, Reston, VA.
Mallin, M.A. 2000. Impacts of animal production on rivers and estuaries. Am. Sci. 88(1):26–37.
Morgan, J.N., and R.C. Messenger. 1973. A sequential analysis program for the analysis of nominal scale dependent variables. Survey Res. Center, Inst. for Social Res., Univ. of Michigan, Ann Arbor.
National Research Council. 1998. Nutrient requirements of swine. Natl. Academy Press, Washington, DC.
Nolan, B.T., B.C. Ruddy, K.J. Hitt, and D.R. Helsel. 1997. Risk of nitrate in groundwaters of the United States—A national perspective. Environ. Sci. Technol. 31:2229–2236.
Qian, S.S., and C.W. Anderson. 1999. Exploring factors controlling the variability of pesticide concentrations in the Willamette River Basin using tree-based models. Environ. Sci. Technol. 33:3332–3340.
Piper, A.M. 1944. A graphic procedure in the geochemical interpretation of water-analyses. Am. Geophys. Union Trans. 25:914–923.
Puckett, L.J., T.K. Cowdery, D.L. Lorenz, and J.D. Stoner. 1999. Estimation of nitrate contamination of an agro-ecosystem outwash aquifer using a nitrogen mass-balance approach. J. Environ. Qual. 28:2015–2025.
Robertson, D.M., D.A. Saad, and A.M. Wieben. 2001. An alternative regionalization scheme for defining nutrient criteria for rivers and streams. Water-Resour. Inventory Rep. 01-4073. United States Geol. Survey, Reston, VA.
Showers, W.J., D.M. Eisenstein, H. Paerl, and J. Rudek. 1990. Stable isotope tracers of nitrogen sources to the Neuse River, North Carolina. Rep. 253. Water Resour. Res. Inst. of the Univ. of North Carolina, Chapel Hill.
Silva, S.R., C. Kendal, D.H. Wilkinson, C.C. Chang, and R.J. Avanzino. 2000. A new method for collection of nitrate from fresh water and analysis for its nitrogen and oxygen isotopic ratios. J. Hydrol. (Amsterdam) 28:22–36.
Smil, V. 1997. Global population and the nitrogen cycle. Sci. Am. 277:76–81.
Spalding, R.F., and M.E. Exner. 1993. Occurrence of nitrate in groundwater—a review. J. Environ. Qual. 22:392–402.
Spruill, T.B., J.L. Eimers, and A.E. Morey. 1997. Nitrate-nitrogen concentrations in shallow ground water of the Coastal Plain of the Albemarle–Pamlico Drainage Study Unit, North Carolina and Virginia. Factsheet 241-96. United States Geol. Survey, Reston, VA.
Spruill, T.B., D.A. Harned, P.A. Ruhl, J.L. Eimers, D.R. Galeone, G. McMahon, K.E. Smith, and M.D. Woodside. 1998. Water quality in the Albemarle–Pamlico Drainage Basin, North Carolina and Virginia, 1992–95. Circ. 1157. United States Geol. Survey, Reston, VA.
StatSoft. 2001. Electronic statistics textbook. Available online at http://statsoftinc.com/textbook/stathome.html (verified 16 May 2001). StatSoft, Tulsa, OK.
Steinhorst, R.K., and R.E. Williams. 1985. Discrimination of groundwater sources using cluster analysis, MANOVA, canonical analysis and discriminant analysis. Water Resour. Res. 21:1149–1156.
Stiff, H.A., Jr. 1951. The interpretation of chemical water analysis by means of patterns. J. Petrol. Technol. 3:15–16.
Therneau, T.M., and E.J. Atkinson. 1997. An introduction to recursive partitioning using the RPART routines. Technical Report. Mayo Foundation, Rochester, MN.
United States Census Bureau. 2001. County population estimates for July 1, 1999, and population change from April 1, 1990 to July 1, 1999. Available online at http://www.census.gov/population/estimates/county/co-99-2/99C2_37.txt (verified 16 May 2002). United States Census Bureau, Washington, DC.
United States Department of Health, Education, and Welfare. 1962. Drinking water standards. Revised. Public Health Serv. Publ. 956. United States Department of Health, Education, and Welfare, Washington, DC.
USEPA. 2001. National primary drinking water standards. EPA 816-F-01-007. USEPA, Washington, DC.
Vitousek, P.M., J.D. Aber, R.W. Howarth, G.E. Likens, P.A. Matson, D.W. Schindler, W.H. Schlesinger, and D.G. Tilman. 1997. Human alteration of the global nitrogen cycle: Sources and consequences. Ecol. Applic. 7:737–750.
Wade, H., C. Bailey, J. Padmore, K. Rudo, B. Williams, and A. York. 1997. The interagency pesticide study of the impact of pesticide use on ground water in North Carolina. North Carolina Dep. of Agric., Raleigh.
Whittemore, D.O., and L.M. Pollock. 1979. Determination of salinity sources in water resources of Kansas by minor alkali and halide chemistry. Kansas Water Resour. Res. Inst., Lawrence.
Wilhelm, S.R., S.L. Schiff, and W.D. Robertson. 1994. Chemical fate and transport in a domestic septic system: Unsaturated and saturated zone geochemistry. Environ. Toxicol. Chem. 13:193–203.
Wilkinson, L. 2000. Classification and regression trees in SYSTAT10. Vol. I. SPSS, Chicago.
Winter, T.C., J.W. LaBaugh, and D.O. Rosenberry. 1988. The design use of a hydraulic potenitameter for direct measurement of differences in hydraulic head between ground water and surface water. Limnol. Oceanogr. 33:1209–1214.
Zublena, J.P., J.V. Baird, and J.P. Lilly. 1991. Soilfacts. Nutrient content of fertilizer and organic materials. Publ. AG-439-18. North Carolina Coop. Ext. Serv., Raleigh.
Zublena, J.P., J.C. Barker, and T.A. Carter. 1993a. Soilfacts. Poultry manure as a fertilizer source. Publ. AG-439-5. North Carolina Coop. Ext. Serv., Raleigh.
Zublena, J.P., J.C. Barker, J.W. Parker, and C.M. Stanislaw. 1993b. Soilfacts. Swine manure as a fertilizer source. Publ. AG-439-4. North Carolina Coop. Ext. Serv., Raleigh.