U.S. Department of the Interior
U.S. Geological Survey

Factors Related to Well Yield in the Fractured-Bedrock Aquifer of New Hampshire

By Richard Bridge Moore, Gregory E. Schwarz, Stewart F. Clark, Jr., Gregory J. Walsh, and James R. Degnan

U.S. Geological Survey Professional Paper 1660

In cooperation with the
New Hampshire Department of Environmental Services

U.S. Department of the Interior

Gale A. Norton, Secretary

U.S. Geological Survey

Charles G. Groat, Director

2002

For sale by
U.S. Geological Survey
Information Services
Box 25286, Federal Center
Denver, CO 80225

Library of Congress Cataloging-in-Publication Data

Factors related to well yield in the fractured-bedrock aquifer of New Hampshire / by Richard Bridge Moore ... [et.al.].
p. cm. -- (U.S. Geological Survey professional paper ; 1660)
Includes bibliographical references.
ISBN 0-607-98453-8 (alk. paper)
1. Groundwater--New Hampshire. 2. Aquifers--New Hampshire. 3. Wells--New Hampshire. I. Moore, Richard Bridge. II. Series.

GB1025.N45 F23 2002
553.7'9'09742--dc21 2002020252

CONTENTS

FIGURES

PLATES

TABLES

CONVERSION FACTORS, ABBREVIATIONS, AND VERTICAL DATUM

Multiply	By	To obtain
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer
square mile (mi^2)	2.590	square kilometer
gallons per minute (gal/min)	0.06309	liter per second
gallons per minute per foot (gal/min/ft)	0.2070	liter per second per meter

Vertical Datum: In this report “sea level” refers to the National Geodetic Vertical Datum of 1929 (NGVD of 1929)—a geodetic datum derived from a general adjustment of the first-order level nets of both the United States and Canada, formerly called Sea Level Datum of 1929.

Factors Related to Well Yield in the Fractured-Bedrock Aquifer of New Hampshire

By Richard Bridge Moore, Gregory E. Schwarz, Stewart F. Clark, Jr., Gregory J. Walsh, *and* James R. Degnan

Abstract

The New Hampshire Bedrock Aquifer Assessment was designed to provide information that can be used by communities, industry, professional consultants, and other interests to evaluate the ground-water development potential of the fractured-bedrock aquifer in the State. The assessment was done at statewide, regional, and well field scales to identify relations that potentially could increase the success in locating high-yield water supplies in the fractured-bedrock aquifer. statewide, data were collected for well construction and yield information, bedrock lithology, surficial geology, lineaments, topography, and various derivatives of these basic data sets. Regionally, geologic, fracture, and lineament data were collected for the Pinardville and Windham quadrangles in New Hampshire. The regional scale of the study examined the degree to which predictive well-yield relations, developed as part of the statewide reconnaissance investigation, could be improved by use of quadrangle-scale geologic mapping.

Beginning in 1984, water-well contractors in the State were required to report detailed information on newly constructed wells to the New Hampshire Department of Environmental Services (NHDES). The reports contain basic data on well construction, including six characteristics used in this study—well yield, well depth, well use, method of construction, date drilled, and depth to bedrock (or length of casing). The NHDES has determined accurate georeferenced locations for more than 20,000 wells reported since 1984. The availability of this large data set provided an opportunity for a statistical analysis of bedrock-well yields. Well yields in the database ranged from zero to greater than 500 gallons per minute (gal/min).

Multivariate regression was used as the primary statistical method of analysis because it is the most efficient tool for predicting a single variable with many potentially independent variables. The dependent variable that was explored in this study was the natural logarithm (ln) of the reported well yield. One complication with using well yield as a dependent variable is that yield also is a function of demand. An innovative statistical technique that involves the use of instrumental variables was implemented to compensate for the effect of demand on well yield.

Results of the multivariate-regression model show that a variety of factors are either positively or negatively related to well yields. Using instrumental variables, well depth is positively related to total well yield. Other factors that were found to be positively related to well yield include (1) distance to the nearest waterbody; (2) size of the drainage area upgradient of a well; (3) well location in swales or valley bottoms in the Massabesic Gneiss Complex and Breakfast Hill Granite; (4) well proximity to lineaments, identified using high-altitude (1:80,000-scale) aerial photography, which are correlated with the primary fracture direction (regional analysis); (5) use of a cable tool rig for well drilling; and (6) wells drilled for commercial or public supply. Factors negatively

related to well yields include sites underlain by foliated plutons, sites on steep slopes sites at high elevations, and sites on hilltops. Additionally, seven detailed geologic map units, identified during the detailed geologic mapping of the Pinardville and Windham quadrangles, were found to be positively or negatively related to well yields. Twenty-four geologic map units, depicted on the Bedrock Geologic Map of New Hampshire, also were found to be positively or negatively related to well yields.

Maps or geographic information system (GIS) data sets identifying areas of various yield probabilities clearly display model results. Probability criteria developed in this investigation can be used to select areas where other techniques, such as geophysical techniques, can be applied to more closely identify potential drilling sites for high-yielding (greater than 40 gal/min) bedrock wells.

To measure the added value of field-based methods for well-yield-probability forecasting, the model was run with and without the variables developed from the detailed geologic mapping for the Pinardville and Windham quadrangles. Four probability maps were produced for the two quadrangles, one with and one without the added variables. These maps clearly demonstrate the advantage of detailed geologic mapping when prospecting for new ground-water supplies in the fractured-bedrock aquifer of New Hampshire.

INTRODUCTION

Considerable population growth has resulted in New Hampshire in the last 40-50 years. This growth has led to a need for additional water resources in many communities. Many of these communities have limited sand and gravel aquifers, which generally are the most favorable aquifers for constructing high-yield wells in New Hampshire. Additional water resources can be found in the fractured-bedrock aquifer, which generally produce low yields to wells (a few gallons per minute). However, through intensive site analyses, high-yielding zones can be located in the aquifer. Throughout New Hampshire, the bedrock is crystalline. Fractured-bedrock aquifers in crystalline bedrock are among the least understood and quantified ground-water resources in the Nation. A major problem in evaluating ground-water availability in fractured bedrock is its extreme variability in water-bearing properties.

The New Hampshire Bedrock Aquifer Assessment was done by the U.S. Geological Survey (USGS), in cooperation with the New Hampshire Department of Environmental Services (NHDES), Water Division, to provide information that can be used by communities, industry, professional consultants, and other interests to evaluate the ground-water development potential of the fractured-bedrock aquifer. This study was done at three scales—statewide, regional, and local—to quantify relations among well yields and bedrock data, physiographic setting, remotely sensed lineaments, and well characteristics in New Hampshire. The statewide scale was designed as a reconnaissance-level statistical analysis of the relations between bedrock well yield and bedrock type, lineament characteristics, topography, and other well-site and construction characteristics. The regional-scale analysis was similar, except that field-based geologic-data collection and additional lineament data were added to assess what effect quadrangle-scale (1:24,000) data had on well-yield relations. Local assessments (usually at a well-field scale) were done by use of geophysical tools to identify the location and orientation of discrete zones of fracture in the bedrock that are referred to as "fracture zones" in this report.

Purpose and Scope

This report presents results of statistical analyses conducted for the statewide and regional scale investigations of the bedrock-aquifer assessment. Results of the well-field assessments are presented in a companion report by Degnan and others (2001).

The NHDES has a large data set of bedrock-well locations, yields, and other characteristics that are adequate for an exploratory, statistical analysis of hydrogeologic and construction factors associated with well yields. There are many possible factors (site and well characteristics) that relate to high-yield well sites. For the statewide analysis, available well data and remotely sensed data were compiled. For the regional analysis, fracture data and general geologic and hydrogeologic information were collected and compiled for two 1:24,000-scale New Hampshire quadrangles. The purpose of the more intensive regional-scale compilation was to determine the degree to which predictive well-yield relations, developed as part of the statewide reconnaissance investigation, could be improved with quadrangle-scale geologic mapping. Data were collected through bedrock mapping and supplemental hydrogeologic observation. The Pinardville and Windham quadrangles were chosen for regional analysis because of the large number of wells located in each quadrangle and the variety of geologic settings in the quadrangles.

This study was not intended to explain the occurrence and nature of water-bearing bedrock fractures in New Hampshire, nor the nature of water movement in the fracture system. Understanding the occurrence and distribution of fractures in crystalline bedrock and associated water movement are being intensively evaluated at research sites (Shapiro and Hsieh, 1991) and cannot be determined statewide or regionally with the present state of the science. The statistical identification of hydrologic and other factors that are related to bedrock well yields statewide and regionally, will, however, provide information to further the understanding of these processes.

Previous Investigations

Large-scale analyses of bedrock-aquifer yields are few, and especially are limited in crystalline bedrock settings in the northeastern United States. Investigators have successfully related geohydrologic factors, particularly lineaments, to well yields in the carbonate rocks of Pennsylvania (for example, Knopman, 1990; Siddiqui and Parizak, 1971). One of the largest studies (Daniel, 1989; Daniel and others, 1983) assessed hydrogeologic characteristics and yields of 6,200 bedrock wells in the Piedmont and Blue Ridge provinces of North Carolina. These investigations found that wells in draws or valleys have average yields three times those of wells on hills and ridges. Wells in the most productive hydrogeologic units had average yields twice those of wells in the least-productive units. One of the investigations, directly applied to fractured-bedrock aquifers in the Northeast, was performed by Mabee and others (1994) on a 17-mi^2 island on the Maine coast. This study examined 35 bedrock wells and found that aquifer transmissivity, normalized by well depth, was positively related to well proximity to fracture-correlated lineaments, which are lineaments trending in the same direction as fractures observed in the bedrock. A number of these studies (table 1) measured the degree to which wells near lineaments represented a population of high yields.

Table 1. Summary of well yield and lineament investigations considered for this study

[>, greater than; NHDES, New Hampshire Department of Environmental Services; NA, not available]

Study/imagery used	Region/units tested	Measure of well productivity	Total number of wells	Number of wells near lineaments	Number of wells beyond buffer zone around lineaments	Confidence the samples represent, two separate populations (in percent)
Siddiqui (1969); Siddiqui and Parizek (1971 and 1974) Low-altitude aerial photography	Pennsylvania carbonate rocks	Normalized transmissivity	45 80	18 53	27	>99
Mabee and others (1994) High-altitude areal photography and side-looking radar	Maine metasedimentary	Normalized transmissivity	35	20	15	30
Mabee and others (1994) High-altitude areal photography and side-looking radar	Maine “Fracture-correlated” lineaments (involving extensive additional fracture-fabric field work)	Normalized transmissivity	35	7	28	89
Preliminary report of present investigation (Moore and others, 1998) Low- and high-altitude areal photography	Plates 1-6, covering southern New Hampshire, “migmatites” and unfoliated plutons (involving no field work beyond NHDES locating wells)	Reported yield	11,212	NA	NA	99.9

Most of the investigations mentioned in table 1 used lineament analysis to identify potential bedrock-fracture zones or bedrock-solution-channel enhancements in carbonate bedrock. Numerous investigations refer to lineament analysis as "fracture-trace analysis"; however, field observations of fractures is desirable for fracture-trace analysis. A lineament is defined as a linear pattern, seen on aerial photographs and other remotely sensed imagery, that meet established criteria for features that may be the result of underlying zones of fractured bedrock (Clark, Moore, and others, 1996). Despite the use of rigid criteria, any given lineament may not be underlain by a fracture zone. Anthropogenic remnants such as an abandoned woods road, rail lines, or right-of-ways, could be identified as a lineament but may not be related to fractures. Lineament data also are limited because of the accuracy of lineament locations, which varies with the scale of the imagery. A lineament analysis, is therefore, a preliminary analysis whereby lineaments are identified on remotely sensed imagery, and can only be confirmed as fracture related through subsequent field observation. Blanchet (1957) described patterns observed on aerial photographs, which were related to fracturing in the Earth's crust. One of the first thorough treatments of lineament analysis was described by Lattman (1958) who provided much of the framework for techniques often applied currently (2001). In New England, many consulting firms commonly use some form of lineament analysis, in combination with other investigative techniques, to locate high-yield well sites in fractured-bedrock aquifers. The specific techniques applied, however, generally are considered to be proprietary and descriptions are often incomplete or not available.

Clark, Moore, and others (1996) describe the lineament-analysis methods used in the New Hampshire Bedrock Assessment. Some of these methods previously were used by Blanchet (1957), Lattman (1958), Brown (1961), Daniel (1989), and Mabee and others, (1994), but no studies integrated results of statewide, regional, and local-scale investigations based on the variety of data compiled for the New Hampshire bedrock-aquifer statistical analysis.

Acknowledgments

The authors thank the individuals who contributed their time and effort offering advise and assistance. Thanks go especially to the following individuals: from New Hampshire Department of Environmental Services, Frederick Chormann, John Cotton, Philip O'Brien, Eugene Boudette, and David Wunsch; from the University of New Hampshire, Mathew Davis, Wallace Bothner, and Lawrence Dingman; from Dartmouth College, John Lyons; from the Maine Geological Survey, Marc Loiselle; from the University of Massachusetts, Stephen Mabee; from the private sector, Gary Smith and Bradford Caswell; and from the U.S. Geological Survey, Thomas Armstrong, Lawrence Drew, James Borchers, Paul Hsieh, Ben Morgan, Jack Schuenemeyer, Allen Shapiro, Charles Daniel and Jeffrey Starn.

HYDROGEOLOGIC SETTING

New Hampshire's landscape is underlain by bedrock, which makes up the crust of the Earth, and by surficial deposits, most of which were left behind by the continental ice sheet during glaciation. The trend of major bedrock units that form the regional structural pattern in New Hampshire generally is north-northeast. The bedrock that underlies New Hampshire is composed of metamorphic and plutonic rocks (Lyons and others, 1997). The geologic history of New Hampshire's bedrock materials, their structure, and the tectonic forces and processes that formed them is part of the history of the Appalachian Mountains. Details of this history can be found in Hatcher and others (1989) and Boudette (1990).

Plutonic rocks, locally common in New Hampshire, are composed of interlocking minerals and have little or no primary pore space. Likewise, primary porosity of the parent materials of metamorphic rock in New Hampshire has been eliminated by metamorphism. Sedimentary and igneous rocks are the parent materials for many of the metamorphic rocks in New Hampshire. These rocks have been subjected to a wide variety of conditions of temperature and pressure, which formed differing assemblages of minerals and textures and eliminated voids by compaction and recrystallization.

Tectonic processes, which accompanied metamorphism and mountain building, folded and faulted New Hampshire's rocks. Tectonic processes created many fractures, faults, and voids and, at the same time, locally sealed some of these structures (Swanson, 1988).

Despite a complex history of sedimentation and multiple events of deformation, metamorphism, and intrusion, New Hampshire's bedrock can be viewed from a broad hydrologic perspective. Water moves through, and is stored in, open fractures. The size, number, distribution, and degree of interconnection of fractures are highly variable; in general, however, fractures are few and, when present, generally are presumed to decrease in size and number with depth. Thus, the overall storage capacity of bedrock is small and tends to decrease with depth (U.S. Geological Survey, 1984, p. 304). Wells that penetrate bedrock commonly yield dependable supplies of water suitable for single-family domestic needs, and, for this purpose, bedrock is a principal aquifer. Zones where bedrock is extensively fractured may yield large quantities of water. Many small water systems that serve residential developments use bedrock wells, and in the past two decades, the application of exploration technology has enabled more municipal water-supply systems to use the bedrock aquifer.

Because ground-water flow processes in fractured crystalline bedrock are complex, developing a more complete understanding of these processes, and methods to characterize them, is a priority of USGS research. The USGS has been conducting investigations at its national fractured-bedrock research site at the Hubbard Brook Experimental Forest in Thorton, N.H., since 1990 (Shapiro and Hsieh, 1991; Hsieh and others, 1993). Research conducted at this site has included developing and utilizing state-of-the-art geologic, geochemical, geophysical, and hydrologic methods to gain a better understanding of ground-water-flow processes in fractured crystalline bedrock.

METHODS AND APPROACH

The approach of this study was to relate bedrock well yields in New Hampshire to factors known to affect well yields in other parts of the country. These factors include bedrock type, lineament characteristics, physiography, and other well-site and construction characteristics on a statewide scale. This analysis was done by use of a multivariate-regression model with instrumental variables. Emphasis was placed on the use of available data and remotely sensed information, with some additional field-data collection involving two geologically diverse 7.5-minute quadrangles. Lineament investigations, as previously discussed, were useful indicators of high-yield wells in crystalline bedrock. Other geologic and site characteristics that were significantly related to well yield also were used in a predictive mode to identify areas with various probabilities of obtaining a selected yield. The approach incorporates the use of statewide and regional information to evaluate the added benefit of additional data collection. The following sections describe the sources and treatment of data and development of the statistical approach.

In investigating factors that relate to high-yield well sites, many possible factors (referred to as variables in this report) were examined. Statewide data sets were compiled from existing sources or were created for this study. The primary information that was analyzed included well information, bedrock lithology, surficial geology, lineaments, topography, and various derivatives of information in those data sets. In addition, data were collected at a more detailed scale for the areas contained in the Pinardville and Windham quadrangles, New Hampshire (fig. 1). These data are discussed in detail in the following paragraphs.

The relation of multiple variables to bedrock-well yields initially was examined by simple regression techniques to compare preliminary analyses with other investigations and to help streamline more detailed regression-model analyses involving the use of instrumental variables. The relation of selected individual variables to yields also was examined by simple, bivariate, statistical analyses.

Regression was selected as the primary method of analysis. As such, the established model is a stochastic (statistical) model rather than a process-oriented geologic model. A process-oriented model (modeling flow in known fractures) was not possible because the necessary detailed geologic information, such as a statewide data set on fractures, is not available. Variables examined included categorical variables (such as bedrock type), as well as continuous numerical variables such as the slope of the land surface. A nested model was used, where detailed geologic data from the quadrangles were applied

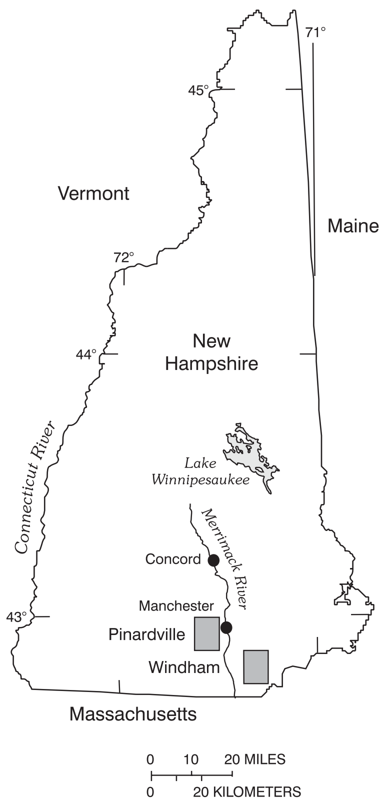

Figure 1. Location of the Pinardville and Windham 1:24,000-scale quadrangles in southern New Hampshire where additional geohydrologic, fracture, and lineament data were collected.

where available or nested in the statewide model. The model was used to evaluate the simultaneous effect of many variables and the interaction of those variables, and provided confidence intervals. These intervals can be used to predict the probabilities of obtaining a given yield given different site and well characteristics.

Descriptions and Sources of Data

This study required the development of geographic information system (GIS) compatible data sets by use of ARC/INFO. All data sets used were the most detailed available statewide. Many variables were considered, however, only 43 were found to be significant and are included in the final regression model.

Well Information

Detailed information on water wells constructed since 1984 is collected and maintained in a computer database by NHDES. Many of these wells have been field-inventoried to obtain accurate locations. Wells are categorized as to whether they are used for domestic, public supply, commercial, or other uses. Nearly 98 percent of the all wells in the database were drilled for domestic use. Basic well data are reported by drillers and include six characteristics used in this study—well yield, well depth, well use, method of construction, date of construction, and depth to bedrock (or length of casing). Well diameter was not included in this database. However, nearly all of the domestic wells are 6 inches in diameter. A georeferenced digital database is maintained by the State where location coordinates were determined either by plotting and digitizing from a map at 1:24,000 or 1:25,000 scale or by differentially corrected Global Positioning System (GPS) measurement. Well locations are accurate to ±100 ft.

The well data set collected by NHDES and used in this study contains records of 20,308 wells drilled from 1984 to 1998, and subsequently field located. A randomly selected subset (20 percent or 4,050 wells) of the available well data were reserved for verification tests of the model. The availability of this large data set provided an opportunity for statistical analysis of variables considered to be related to well yields. Specific components of the database used in this study are described in the following paragraph with respect to their use for statistical analysis.

Well yields in the database ranged from zero to greater than 500 gal/min. Yields are reported by well drillers and are determined by a variety of methods at the time the well is drilled. Yields generally are determined as the rate of water that can be airlifted on a continuous, short-term (generally tens of minutes) basis. Low yields may be accurately measured by timing the pumpage of water into a known volume. Moderate- to high-reported yields often are quantified by discharging water through a rated channel and are likely to be rough approximations of the true yield. Such methods introduce an unquantifiable variance in measured yields but are acceptable for statewide or regional evaluations involving large numbers of wells. Yields equal to or greater than 40 gal/min were considered high in this investigation. High-yielding wells constitute nearly 10 percent of the well population (fig. 2). The yields of public-supply wells typically are quantified by an aquifer test of hours to days in duration and represent a more rigorous estimate of yield than that reported for domestic wells by the drillers. The dependent variable that was statistically explored in this study was the natural log of the reported well yield. Other possible dependent variables considered were the yield and yield per foot of open hole; however, the natural log of well yield was selected because the population of well yields is log normally distributed.

A problem in using well yield as the dependent variable in a multivariate-regression analysis is that yield is a function of a well owner's water needs. Generally, a well only is drilled as deep as is needed to meet a specific targeted yield. Thus, the reported yield is not the maximum potential yield for that site. As well yield is, in part, a function of demand, an advanced statistical technique involving instrumental variables must be used for a more accurate estimate of yield.

The year each well was drilled can be included as a surrogate for demand in the model because demand for water (and the depths to which well owners have been willing to pay to drill to obtain more water) has increased in New Hampshire. The effect of time was observed using data from the Pinardville quadrangle (Lawrence Drew, U.S. Geological Survey, oral commun., 2000). Drillers confirmed that well depth increased to meet this demand (Terry Swain, Capital Well Co., oral commun., 2000).

Depth, reported by the drillers, likely is accurate to the nearest foot. Depth also is a function of demand and is inversely related to well yield. Wells are drilled deeper at low-yield sites and drilling is stopped when a desired yield is reached. This action makes depth an "endogenous variable," and, for this reason, depth is inversely related to well yield.

Data on the method of well construction essentially are limited to two general categories; cable tool and rotary. Bedrock wells drilled with a cable-tool drill rig can be identified in the NHDES database, but all other methods, such as the various forms of rotary drilling, were not differentiated. As cable-tool construction is different from the other techniques, it is reasonable to test the effects of cable tool in comparison to other methods of construction in the regression model. With cable-tool construction, the drill bit is pounded into the bedrock perhaps opening up or enhancing fractures. An indicator variable was used to identify whether or not each well was constructed by use of a cable-tool drill rig.

Bedrock Lithology

Bedrock lithology was assigned using a digital data set of the Bedrock Geologic Map of New Hampshire (Lyons and others, 1997). This data set contains 174 mapped bedrock units and was compiled at a scale of 1:250,000.

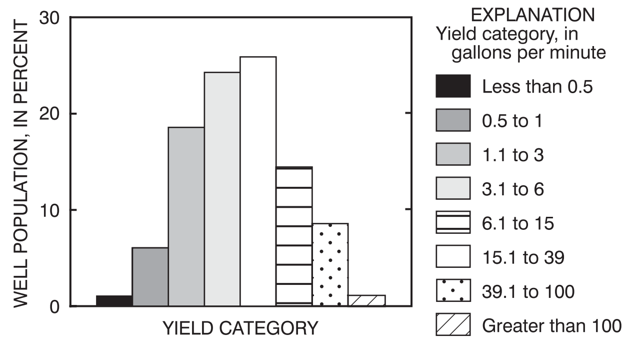

Figure 2. Percent of statewide bedrock well yields in New Hampshire.

In addition to the 174 mapped bedrock units, 7 generalized categories, covering all the rocks of New Hampshire, were created with input from Dr. Wallace Bothner (University of New Hampshire, written commun., 1998). Variables identifying these groupings were created to broadly categorize similar rock types. Some mapped geologic units depicted on the State bedrock map have limited areal extent, yet have properties such as origin, texture, and degree of metamorphism common to other mapped units. Grouping these lithologies together into common categories allows for statistical analysis of areas that would otherwise have had too few well-data points to be included individually as independent variables. These seven major lithologic units with similar lithologic characteristics are (1) nonfoliated to weakly foliated igneous rocks, (2) well-foliated igneous rock (well-foliated rocks of the New Hampshire Plutonic Suite and Oliverian Plutonic Suite), (3) fine-grained, tightly bonded metasedimentary rocks, (4) fine to medium-grained granular-metasedimentary rocks, (5) gray-black graphitic micaceous metapelite locally interbedded with calcareous or siliceous rock, (6) volcanic rock and associated metapelite, and (7) rocks of the Massabesic Gneiss Complex and Breakfast Hill Granite (fig. 3).

Lineaments

A statewide data set of lineaments was created for this project following the methods presented in Clark, Moore, and others (1996). The imagery used came from four remotely sensed platforms. These platforms included Landsat imagery (1:1,000,000 enlarged to 1:250,000), side-looking airborne radar (SLAR) (1:250,000), high-altitude aerial photography (approximately 1:80,000), and low-altitude

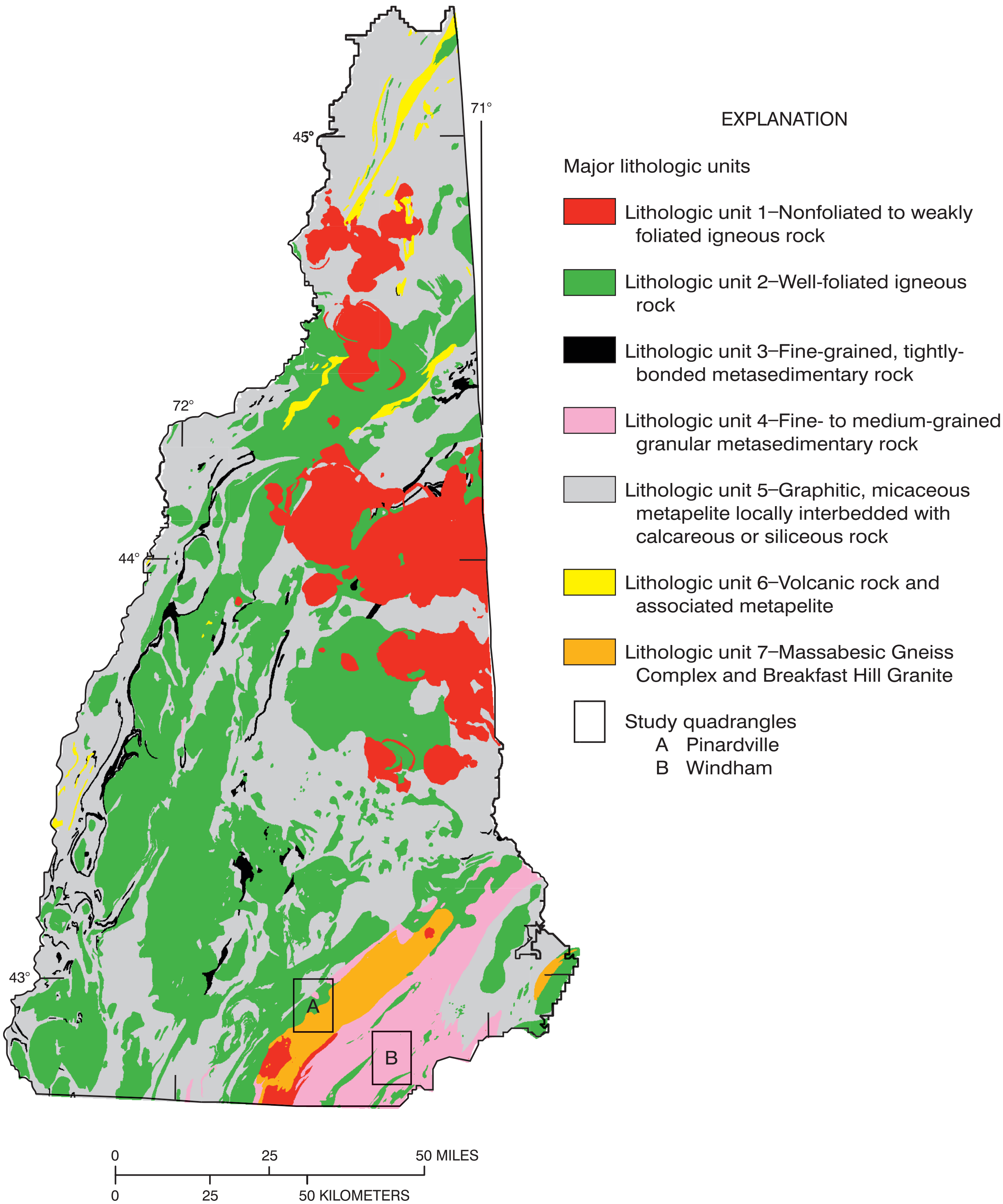

Figure 3. Major lithologic units in New Hampshire adapted from Lyons and others (1997).

photography (approximately 1:40,000, enlarged to 1:20,000). In addition to the imagery mentioned above, the regional-scale analysis included color-infrared imagery (approximately 1:40,000) and topography from standard (7.5 minute) topographic maps (1:24,000). Data at different scales were analyzed by different analysts in order to maximize the reproducibility of lineament identifications (Clark, Moore, and others, 1996). This analysis was done by selecting just those lineaments identified by two different observers. The imagery was examined and lineaments were recorded on mylar overlays by the observers working independently. Results were compared and placed in the following three categories: (1) blind comparisons—matching lineaments drawn by independent observers; (2) confirmed lineaments—lineaments drawn by one observer and confirmed by another observer; and (3) rejected lineaments—lineaments drawn by one observer and not confirmed by another observer. All blind or confirmed lineaments were digitized into a GIS database for use in the statistical analysis. These features were either plotted on scale-stable base maps that were later mounted on a digitizing tablet and digitized, or were plotted directly on imagery overlays that were scanned later and rectified to allow on-screen digitizing.

Lineament data are limited because (1) the accuracy of lineament locations varies with the source and scale of the imagery, (2) lineaments can be caused by cultural rather than geologic features, (3) lineaments related to bedrock features are most easily observed in areas where the surficial material is thin (areas where overburden is thin are known to have low-yielding bedrock), and (4) lineaments theoretically are restricted to steeply dipping fractures or fracture zones that produce straight-line intersections with the earth surface. As lineaments are more readily found in the hilly shallow-to-bedrock areas, establishing a relation to well yield is difficult because of inaccuracies in the georeferenced locations of the wells and the lineaments, and because these shallow-to-bedrock areas also are associated with relatively low yields.

For each well site, distance to lineaments and lineament characteristics were compiled for use in the statistical analyses. Distance from wells to lineaments was coded by identifying whether or not the wells were in a "buffer zone" surrounding each lineament. A buffer zone of 100 ft was selected by trial-and-error, considering the accuracy of the well and lineament locations. The accuracy of lineament locations is approximately ±80 for aerial photography and ±150 ft for the Landsat imagery. Well locations are estimated to have an accuracy of ±100 ft. Selecting a buffer zone less than 100 ft would, thus, have little or no meaning. The buffer zone selected (100 ft) is consistent with the findings of Mabee (1992, p. 106), who selected (by trial-and-error) an optimal buffer zone, for statistical analysis, of 98 ft (30 m).

Lineaments identified from SLAR were excluded from the analysis because of difficulties in positioning the lineaments even to an accuracy of ±100 ft. The SLAR data, compiled on 1:250,000 scale maps, were located to approximately ±200 ft. Additionally, because SLAR relies on shadowing effects, there may be a tendency for the apparent location of the lineament (the shadow) to be off-center from the bottom of the valleys or swales that cause the feature. Linear cultural features, such as road cuts, rail lines, right of ways, and tree lines at the edges of open spaces, can be prominent on SLAR imagery (more so than on other imagery) and could be mistaken for geologic lineaments. The SLAR imagery was carefully examined to avoid misidentification; however, the difficulties in positioning the SLAR lineaments precluded its use.

Additional variables were created to evaluate interactions between lithology and lineaments. Well yields in relation to lineaments or fractures can differ between lithologies because the character of fractures and the extent of the affected area near fracture traces may differ between lithologic units. For example, weathered fracture zones in plutons may contain permeable "grus," a granular fragmental product of weathering of granitic rocks (Bates and Jackson, 1980), but in metasediments may contain clay-rich low-permeability saprolite. The character of fractures and the extent of affected areas near fracture traces also can differ between bedrock types. In plutonic rocks, for example, there tend to be more sheeting fractures, which could be hydrologically connected to vertical fracture zones.

The effect of the orientation of the lineaments was examined late in the process of building the regression model to minimize the number of possible variables. The strike of the nearest lineament within 100 ft of each well was computed and binned into eighteen 10-degree-angle categories. Eighteen angle categories multiplied by numerous lithologies would result in an inordinate number of variables with the likely possibility of spurious results.

A lineament-density database for New Hampshire was created by identifying the density of lineaments within a 1,000-foot radius around each well for each lineament platform (Clark, Moore, and others, 1996, p. 10, plates 1-11. Plate areas 1-10 include all of New Hampshire south of 44 degrees latitude and plate 11 is just to the north and includes most of the eastern part of the White Mountains). Lineament density was equal to the total length of all lineaments within 1,000 ft of each well divided by the area of this circle. Contrary to expectations, lineament density was found to be negatively correlated to well yield in these preliminary analyses. Thus, areas with high lineament densities are most likely in areas of thin overburden, which are inversely related to well yield. (Fracture zones beneath thick overburden are less apt to be observed). Because of this negative relation, and the availability of better predictors of areas that have shallow depths to bedrock, lineament density was not used in further well-yield analyses. Similarly, the distances from the wells to where lineaments intersect one another were examined for five test quadrangles. Unlike positive relations found elsewhere in carbonate bedrock, no significant relations were found in the New Hampshire setting; therefore, intersections were not used in further analyses.

Surficial Material

Surficial material such as stratified-drift aquifers possibly affect bedrock well yield through (1) increased permeability or the ability to supply water to the bedrock aquifer, (2) saturated overburden at the bedrock interface, and (3) greater occurrence of stratified-drift aquifers in valleys. New Hampshire lacks a statewide GIS data set of overburden types; however, a statewide stratified-drift aquifer data set is available and can be applied to the regression model by use of a dichotomous indicator (0 or 1 binary) variable. Overburden type was obtained from maps of the extent of stratified-drift aquifers in New Hampshire, which were mapped as part of a statewide series of investigations (Medalie and Moore, 1995; Moore and others, 1999). Areas outside the mapped stratified-drift aquifers can be classified as till and (or) bedrock. An indicator data set (1 or 0) was used to identify whether a well penetrates through stratified drift or not. No further differentiation of surficial materials is available on a statewide basis.

The thickness of non-stratified-drift deposits is not mapped statewide but can be determined or estimated at individual well sites. Most well drillers' reports contain the depth to bedrock determined during drilling. Where these data are incomplete, the length of casing minus 12 ft can be used as a surrogate for depth of overburden. Twelve feet is the median depth that casings are set into bedrock based on 18,330 wells statewide. Drillers prefer to end the casing well in competent bedrock to avoid the upper few feet, which may be weathered or fractured. Because there is no statewide GIS data set of overburden thickness, this variable could not be used in a predictive mode.

Topographic Settings and Characteristics

Data were compiled to provide various sets of topographic and physiographic information that may be associated with bedrock well-yield characteristics. Topographic settings can be derived from analysis of USGS Digital Elevation Models (DEMs). DEMs consist of a sampled array of elevations for a number of ground positions at regularly spaced intervals. The DEM data for 7.5-minute units correspond to the USGS 1:24,000 and 1:25,000-scale topographic quadrangle map series. Each 7.5-minute DEM is based on 30 x 30-m data spacing with the Universal Transverse Mercator (UTM) projection. The 1-degree DEMs, with 3 x 3-arc-second data spacing (or roughly 100-m spacing) provides coverage in 1 x 1-degree blocks. The 1-degree DEMs also are referred to as 1:250,000-scale DEM data. Land-surface elevations of wells determined from the 7.5-minute DEMs were included in the model. Lower well yields are expected in high elevations as a result of greater depths to water, less saturated thicknesses, and(or) the predominance of competent rock with few fractures.

Topographic setting also can be analyzed regionally to identify, for example, the effects of major river valleys. This analysis is done by use of 1:250,000 scale DEMs. Effects of more localized features, such as swales or cuts in ridgelines, can be assessed using 1:24,000-scale DEMs. Slope, or the rate of change in elevation surrounding a well, was tested at both scales. Slope at the 1:24,000-scale can be used to compare the change in elevations between neighboring DEM elevation points about 100 ft away from one another. Slope at the 1:250,000-scale DEM is a more regional measurement. Slope expressed as a percent is calculated in ARC/INFO relative to the direction of maximum change in elevation. Assuming similar results to those for North Carolina (Daniel, 1989), slopes or hillsides are expected to have average well yields, and valley bottoms or depressions are expected to have above-average well yields. Wells on hills and ridges are expected to have below-average yields.

Curvature, calculated as the second derivative of the land-surface elevation, provides another quantifiable measure of topographic setting. The curvature function used in this study fits a fourth order polynomial surface to each 3 x 3 block of DEM data points (referred to as GRID cells in ARC/INFO) surrounding and including the central DEM point for the calculated curvature. Negative values indicate that the land surface is concave upward (for example, a swale or valley bottom) and positive values indicate the land surface is concave downward (for example, a hilltop). Again, use of the 1:250,000-scale DEM results in a coarse resolution with a much more generalized regional depiction of the land surface, and cannot be used to detect small features such as depressions in open areas or small gaps in ridges. Curvature at the 1:24,000-scale, however, will not show large physiographic features at the well site. Sites where the topography was concave upward (especially with gentle slopes) represented locations that were eroded more than the surrounding rock and may be underlain by fractures. Conversely, hilltops, which are concave downward, are more resistant to erosion and may be related to a decrease in fracture density and well yield.

Categorical variables were created for combinations of slope and curvature. These variables are "dichotomous," in that they indicate with a 0 or a 1 whether a given condition is met. For example, the quartile of the well population with steepest site slopes was identified as being on hill slopes (dichotomous variable is set equal to 1). The cross between the most negative quartile of the population, for curvature, with the quartile of the population with the most gentle slopes, was used to identify swales or valley bottoms.

Drainage area, upgradient from each well site, was determined by use of the 1:24,000-scale DEM and a computer process known as "flow accumulation." First, isolated low points ("sinks") surrounded by high points are filled in. These points typically represent an artifact of the DEM. Next, the downslope direction between all DEM cells was determined, and the number of cells upgradient of every cell was determined (including that cell). This flow accumulation was done statewide to measure the drainage area of each well site.

The distance of each well to the nearest waterbody was determined by use of the 1:24,000-scale Digital Line Graphs (DLG) of hydrography. For this method, only shorelines and perennial streams were used to compute the distance. Wetlands and intermittent streams were excluded because they are depicted less accurately and consistently on the DLGs than are open perennial waterbodies.

Additional Variables Investigated at the Quadrangle Scale

Additional geologic, fracture, and lineament data were collected for the Pinardville and Windham quadrangles (fig. 1). These quadrangles were selected because they represent different geohydrologic settings, and because these quadrangles contain the largest number of georeferenced bedrock wells anywhere in the State of New Hampshire. Different degrees of metamorphism have affected the rocks present in these quadrangles with much of the Pinardville quadrangle containing the Massabesic Gneiss (migmatitic) Complex. In the Pinardville quadrangle, there are 1,682 bedrock wells with a complete set of well information in the database, and for the Windham quadrangle, there are 1,504 wells. These data provided a large enough data set to evaluate the effects of additional variables, such as fracture-correlated lineaments, lineaments identified from topographic maps and color infrared imagery, and additional variables identified by use of quadrangle-scale geologic mapping.

Topographic lineaments, drawn on a 1:48,000-scale topographic base reduced from the 1:24,000-scale topographic maps, and lineaments drawn on 1:58,000-scale color infrared photographs were added to the lineament database for the Windham and Pinardville quadrangles. Topographic maps are a readily available platform commonly used for lineament identification. Color-infrared photography especially is sensitive to anomalies in vegetation growth, which may be related to underlying fracture zones. Vegetation type and intensity can vary with ground-water recharge and discharge zones. The same procedures described in Clark, Moore, and others (1996) were followed in this process. Color-infrared photographs with overlays were rectified and both sets of lineaments were digitized into GIS data sets for use in the statistical analysis.

Geologic mapping at the 1:24,000 scale in the Pinardville quadrangle (Thomas R. Armstrong and William C. Burton, U.S. Geological Survey, written commun., 2000) and Windham quadrangle (Walsh and Clark, 1999) provided an accurate representation of lithologies and data on brittle fracture and ductile structure orientation for correlation with lineaments. These maps accurately depict geologic units, contacts, fault, axial trace, metamorphic isograd, dike and other features, which were tested relative to well yields. Simplified versions of the maps showing geologic units are shown in figures 4 and 5.

The geologic units mapped at the quadrangle scale include numerous subdivisions of the geologic units shown on the Bedrock Geologic Map of New Hampshire (Lyons and others, 1997). The digital GIS coverage of the State map, used for development of the regression model, was adjusted where contacts mapped at the quadrangle scale separate units shown on the State map.

Other geologic features in the quadrangles similarly were assessed as the lineaments in the regression model. These features include contacts, faults, dikes, and axial planes of folds large enough to be mapped at a scale of 1:24,000. Indicator variables were created to indicate whether a well was within 100 ft of one of these features. Geologic contacts, faults, axial traces, metamorphic isograds, and dikes also were matched with lineaments in the Windham and Pinardville quadrangles. Lineament segments that fall on, or are parallel within 200 ft of these features, were identified and tested to determine if they were associated with high-yield wells.

Information on low-angle jointing was not available statewide but was collected for the two quadrangles. The occurrence of low-angle joints near the anticlinal axis of the Massabesic Gneiss Complex (fig. 4) is a prime example of where these features can be related to well yield. An indicator variable was created to test whether or not wells within 2,000 ft of the mapped axis of this broad anticlinal feature had high yields.

Whether some lineaments are more likely to be related to fractures than others has prompted the search for lineaments that are "fracture correlated." Mabee and others (1994) recognized the need to filter, or reduce, lineament data sets before correlating these remotely sensed features with well yield. Their method for identifying fracture-correlated lineaments is based on identifying fracture domains. Every fracture at a few select outcrops (fracture stations) was sampled and the geographic extent of the fractures was identified by comparing fracture families at each fracture station. Fracture families, as defined by Mabee and Hardcastle (1997, p. 24), are "a set of fractures that have similar orientations." A comparison of studies done at different scales by Mabee and others (1994) and Mabee and Hardcastle (1997) shows that the sizes of fracture domains appear to be related to the spacing of fracture stations. Fracture domains, therefore, appear to be scale-dependent and the domain boundaries can be arbitrary if they cannot be correlated to a specific geologic feature.

In the New Hampshire Bedrock Aquifer Assessment, fracture families were derived using fracture data from the entire mapped area rather than just selected stations. Walsh and Clark (2000) compared mapped fracture orientations in the Windham quadrangle to fracture orientations at selected stations (large outcrops in the Windham quadrangle where they collected additional data). The results show a correlation between station fractures and mapped fractures for the quadrangle. However, about one-ninth of the quadrangle could not be correlated and the mapped fractures surrounding a given fracture station may contain fracture families not represented by the fracture-station data. For the development of the regression model, fracture-correlated lineaments are based on the geographically, more extensively mapped fracture data.

Lineament data were analyzed to see if geologic mapping at the quadrangle scale could identify a set of lineaments with improved correlation to high-yield wells. The lineaments were categorized in the model in the following three ways: (1) unfiltered lineaments that include data from the entire statewide database; (2) filtered, domain-based fracture-correlated lineaments; and (3) filtered, discrete-analysis-based fracture-correlated lineaments (fig. 6).

Domain-based fracture-correlated lineaments were determined by defining fracture families for each of the two quadrangles. A square-cell sampling grid, or domain, was developed so that no cell would have less than five sample points. The optimal grid cell measured 3,300 x 3,300 m. Fracture families for each cell were derived from spatial analysis of mapped fracture data for each of the two quadrangles by plotting frequency-azimuth (rose) diagrams in the Structural Data Integrated System Analyzer software (DAISY 2.19) by Francesco Salvini, Dipartimento

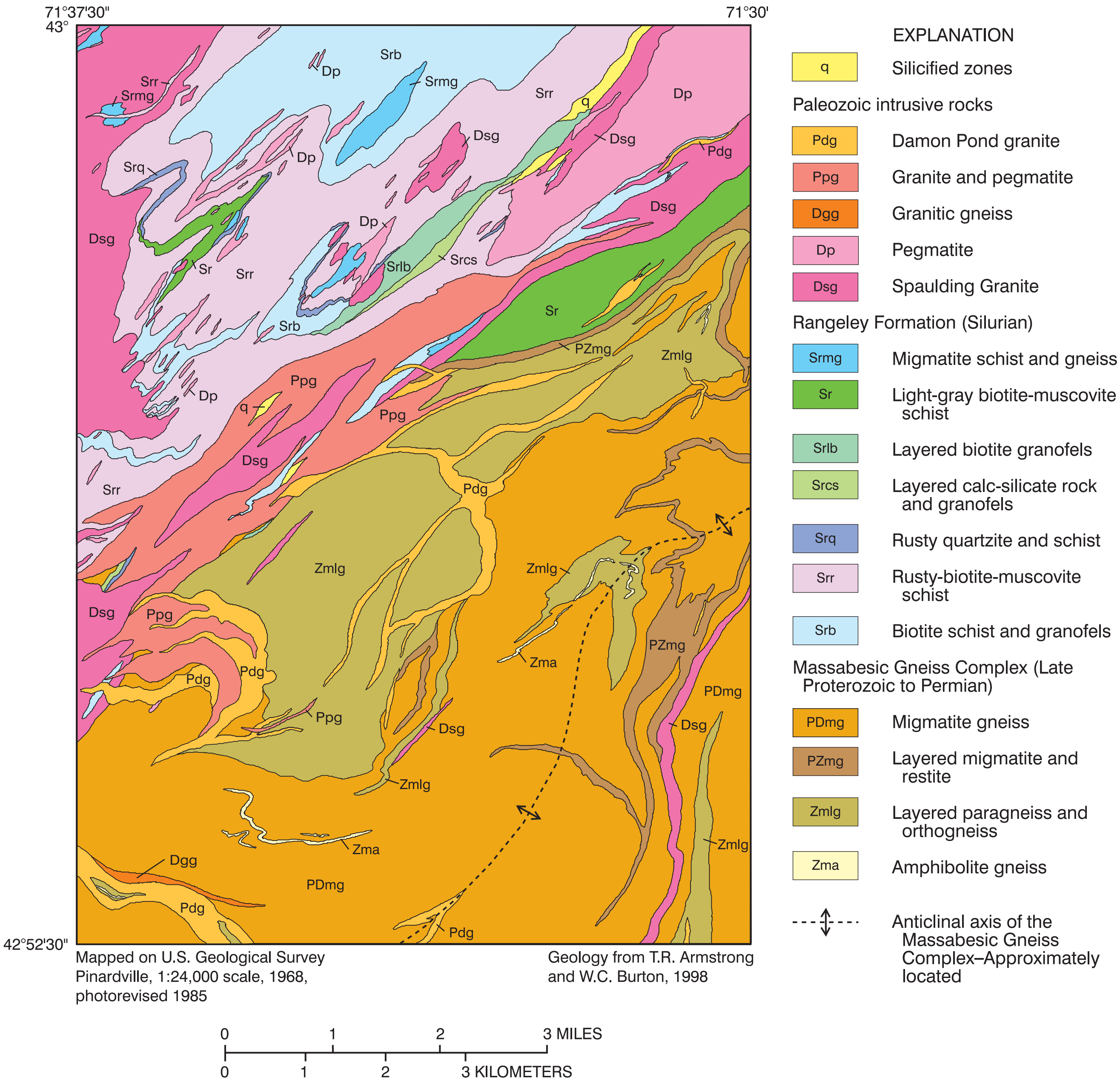

Figure 4. Simplified bedrock geologic map of the Pinardville quadrangle, New Hampshire (Location of quadrangle shown in figure 1.)

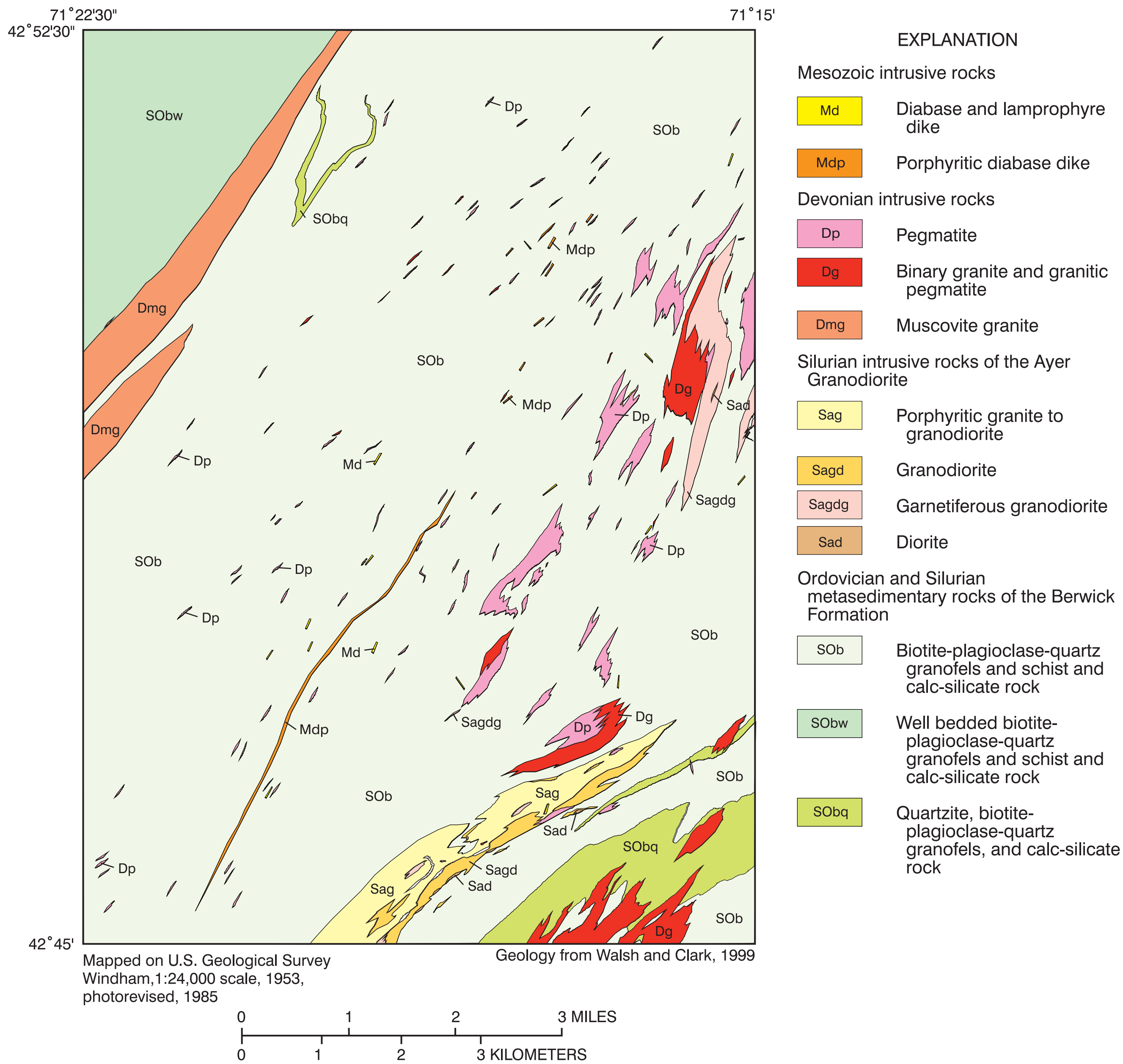

Figure 5. Simplified bedrock geologic map of the Windham quadrangle, New Hampshire (Location of quadrangle shown in figure 1.)

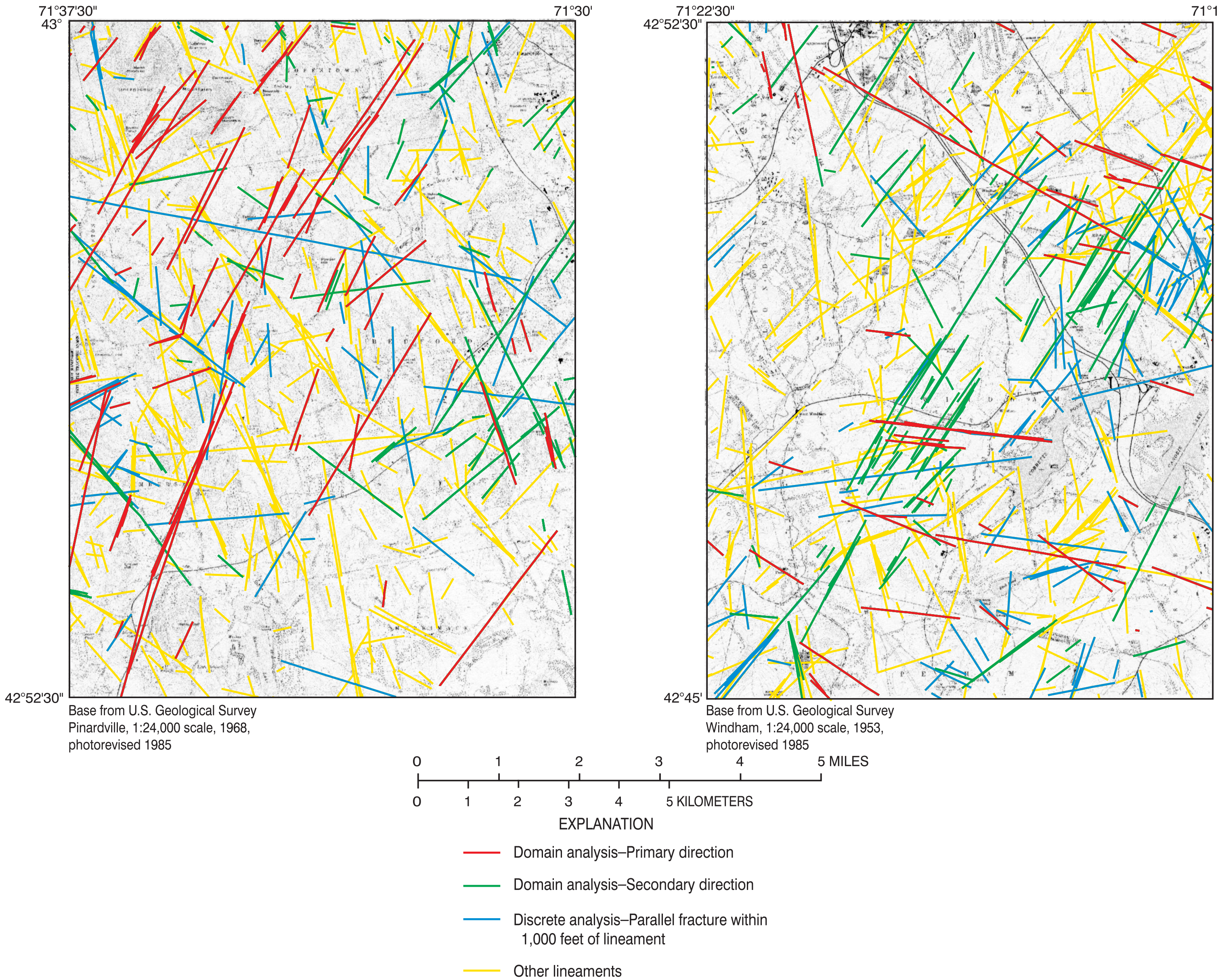

Figure 6. Fracture-correlated lineaments in the Pinardville and Windham quadrangles, New Hampshire.

di Scienze Geologiche, Universita degli Studi di "Roma Tre", Rome, Italy (computer software, 2000). The DAISY software uses a Gaussian curve-fitting routine for determining peaks in directional data (Salvini and others, 1999) that was first described by Wise and others (1985). The rose diagrams included strike data for steeply dipping fractures (dips greater than 45°, after Mabee and Hardcastle, 1997). In this method, primary fracture families on the rose diagrams were defined as the highest peak, and secondary fracture families were defined as normalized peaks between 30 and 99 percent of the highest peak. Lineaments that are in or pass through a given cell, and are parallel to the fracture family for that cell, are identified as fracture correlated. Parallelism is defined by lineament trends that match primary and secondary fracture families within one standard deviation as determined by the Gaussian curve-fitting routine. The term "fracture correlated" in the domain method follows the definition by Mabee and others (1994) in that the correlated lineaments cannot be used to unconditionally identify fractures on the ground, but merely to imply that fractures are likely in the given area. Results from the domain analysis indicate that, by lineament length, 30.5 percent of the lineaments in the two quadrangles correlate with fracture trends, and 14.6 percent correlate with the primary fracture trend. These filtered, domain-based fracture-correlated lineaments were tested in the regression model.

Discrete-analysis-based fracture-correlated lineaments were determined by use of a more rigorous and conservative method than domain analysis (fig. 6). Domain analysis assumes that a statistically identified fracture family is applied to an entire area or domain, and, thus, only can be used to filter groups of lineaments that match the fracture trend in that domain. The domain approach does not address spatial variability in the domain and spatial variability is inherent in quadrangle-scale fracture-trend analysis (Walsh and Clark, 2000). This discrete-analysis method examines each lineament individually, or discretely, by comparing the strike of each lineament in a data set with the strike of nearby steeply dipping planar features or the trend of linear features in a structural geology data set. The discrete method identified lineaments that correlated with many types of structures (including fractures) measured during geologic mapping. The discrete method identified individual lineaments that matched the strike or trend within ±5° (after Mabee and others, 1994) of the structures observed in outcrops up to 1,000 ft (305 m) from a lineament. Lineaments that match these structures are called "structure-correlated lineaments," and may be parallel to ductile or brittle features identified during quadrangle-scale geologic mapping. A subset of these lineaments includes the brittle fracture-correlated lineaments, and this subset was tested for inclusion in the regression model. Results of the discrete method indicate that, by lineament length, 30.9 percent of the lineaments were identified as fracture correlated.

Creation of a Verification Data Set

Twenty percent of the well data were randomly selected statewide and set aside as a verification data set. Sequestered data are used for verification of testing results of the model developed by use of the primary data set. The primary data set contained data from 16,302 wells. The verification data set contained data from 4,050 wells.

Development of the Statewide Regression Model

Multivariate regression was selected as the primary analytical technique because it is the most efficient tool for predicting a single variable (such as the natural log of well yield) with many potentially independent variables. Multivariate-regression analysis, unlike simple bivariate analyses, examines the added contribution of each parameter in explaining the variance of the dependent variable. The regression model was structured to account for drillers targeting well yields to specific demands. For a group of wells drilled to meet similar demands, such as domestic needs, the result is a minimized range of yields. If well depths were random, or if a uniform depth was drilled at each well site, there would be a greater range, or variance, among the well yields.

Because domestic wells make up most of the well-yield data set, the demand criteria by homeowners needs to be factored into the analysis. Domestic needs generally are met by a yield of a few (usually less than 10) gallons per minute, and the wide range of yields is decreased by targeting similar yields. At a potentially high-yield site, drilling typically is stopped at a shallower depth than at an average site, which effectively reduces the number of deep high-yielding wells in that population. At low-yield sites, drilling is continued to a greater depth to obtain a desired yield, which consequently reduced the number of low-yielding shallow wells. Applying an ordinary least squares model to the relation between yield and depth would result in a model that was biased downward; that is, the coefficient of the depth would be lower than its true value in an unbiased data set.

Assuming that drillers are targeting yield, the variable "DEPTH" was "endogenous" (not independent of yield but developed with yield) and correlated with the residual of the yield regression. A regression of yield on depth violates the regression assumption that predictor variables and the error term are independent. To overcome this problem, the statistical technique of using instrumental variables (Davidson and MacKinnon, 1993) was applied in the New Hampshire Bedrock Aquifer Assessment. By this technique, the model evaluates the effects of demand. The bias, caused by the endogenous depth variable, can be corrected if there are variables (instrumental variables) that are correlated with yield demand but uncorrelated with random variation in the physical yield-depth relation. Initially, an attempt was made to include water use in the model because it is related to demand requirements. The water-use variables tested were the indicator variables that identify if the well is a commercial well or a public supply well. Other factors, likely related to well-construction techniques, such as well diameter, however, cause higher well yields in public supply and commercial wells than yields in domestic wells and water-use variables were, thus, not useful. Other more useful instrumental variables were (1) year drilled, (2) well driller, and (3) median household income based on the 1990 census (U.S. Census Bureau digital data, 1990). These variables are correlated with desired yield but are uncorrelated with random variation in the physical yield-depth relation.

Well data from the Pinardville quadrangle (Lawrence Drew, U.S. Geological Survey, oral commun., 2000) indicated that the yields of newly constructed wells, measured by the drillers, increased with time because of an increase in demand for water. Local drillers confirmed that they gradually are drilling new wells deeper to meet this increased demand. Year of drilling can be used as a surrogate for increased demand, and it was introduced into the regression model as an instrumental variable.

A two-stage regression model with instrumental variables was used to control the depth variable. The principle is depicted in figure 7. In this example, only year of drilling is used as an instrumental variable. Conceptually, it is included in the first stage of the model, which predicts lnDEPTH, the natural log of depth. The predicted lnDEPTH then is used in the second stage as a predictor. By excluding year from the second stage and substituting the predicted lnDEPTH from the first stage, a depth coefficient describing the physical relation of depth to yield, rather than the demand relation, is computed (fig. 7). The independent variables in the model are accounted for and the differences from year to year define the lnYIELD to lnDEPTH relation, providing a coefficient for depth that represents the physical relation.

A statistical analysis of well depth by drilling company, for the 25 major drillers in New Hampshire, indicated that 21 drillers tend to drill deeper or shallower than average. As a result, drilling company also was used as an instrumental variable that is correlated with desired yield but uncorrelated with random variation in the physical yield/depth relation.

The spatial distribution of median annual household income, determined from the TIGER data for the 1990 census (U.S. Census Bureau, 1990), also was used as an instrumental variable (fig. 8). Generally, wells in areas of high-median income are assumed to be drilled deeper to meet a somewhat higher demand for water than in areas of low-median

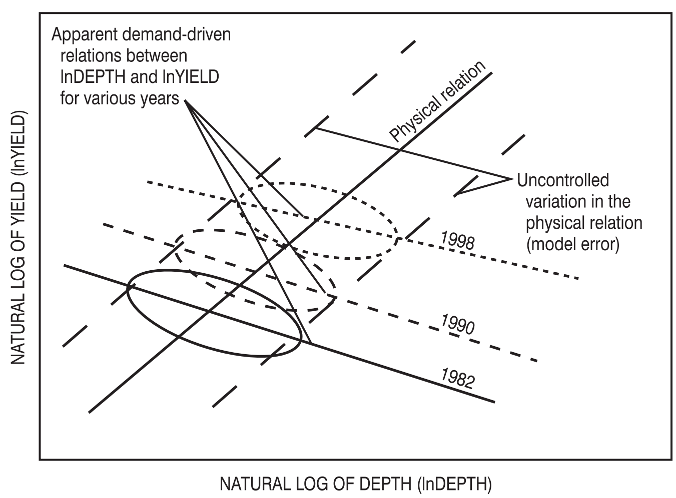

Figure 7. Use of the instrumental variable "YEAR," and a two-stage regression model to separate demand-driven relations from physical relations between well depth (lnDEPTH) and yield (lnYIELD) used in this study of New Hampshire.

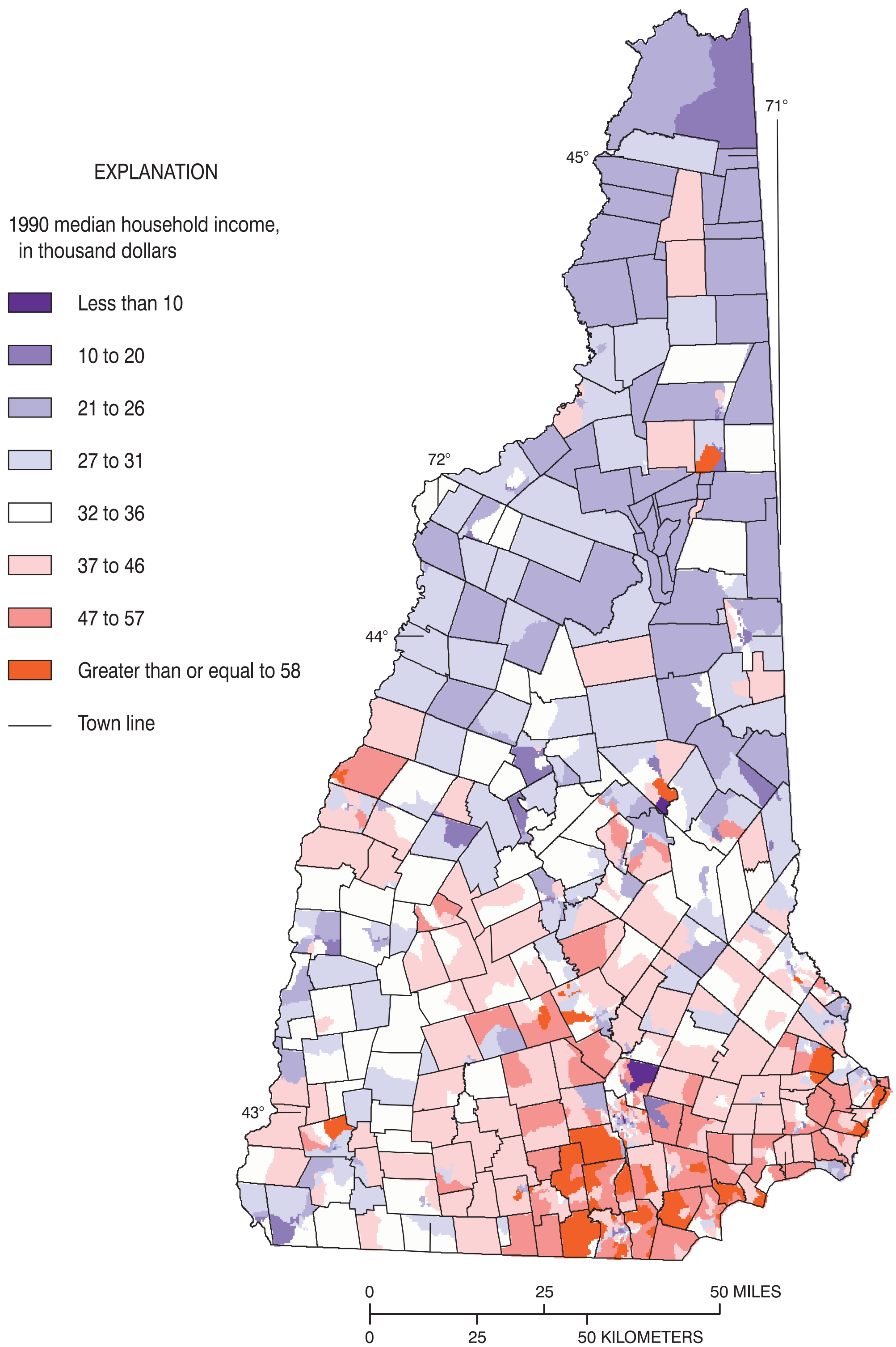

Figure 8. Median household income for New Hampshire using the 1990 U.S. Bureau of Census data. Median income was used as a surrogate demand variable (instrumental variable) in the model.

income. Where census polygons lacked data, because of no population in 1990, the median county income was assigned to the income polygon. (This assignment was done to cover all possible wells, including those wells drilled post 1990.) Each well was assigned a median income based on the polygon in which the well was located and this income was used as another instrumental variable.

Residuals from the two-staged regression model are computed as the difference between actual lnYIELD and predicted lnYIELD from the regression using actual lnDEPTH rather than predicted lnDEPTH. This process was used to essentially estimate the variance of a population of wells that would exist if yield had not been a function of demand. Details on how the model was built are given in appendix A. A list of variables considered and tested for inclusion in the model are given in appendix B.

FACTORS RELATED TO WELL YIELD

Numerous variables were tested for inclusion in the regression model. The strategy for adding or deleting variables during the model construction is outlined in appendix A. All variables included have met a 95-percent confidence level.

Model results show that bedrock well yields are related to a number of physiographic, geologic, socioeconomic, and various other factors. These factors are best illustrated by examining the regression model equation, which relates independent variables to yield. Testing of the model provides confidence in the relations. The model then can be used to forecast and create maps of yield probability.

Regression Model Equation

There are a total of 66 variables used to predict lnDEPTH in the first stage of the model, of which 43 are included in the second stage of the model along with the predicted lnDEPTH variable (table 2). Of these 43 variables, all but the two water-use variables (commercial- and public-supply wells) and the method of construction variable are site characteristics. Many of the variables also are dichotomous-indicator variables and are zero where a given condition is not met. For this reason, at any one location, there is a maximum of 12 site-dependent variables (with any indicator variables equal to 1) and a minimum of 5 site-dependent variables that can apply. For all indicator variables, the magnitudes of the coefficient can be directly compared to one another. Indicator variables with large coefficients (in either a positive or negative direction) have greater effects on lnYIELD than do indicator variables with smaller coefficients. The coefficients of the continuous variables, such as elevation and slope, cannot be directly compared to one another in this manner.

The instrumental variable (two-stage) approach affects the calculated regression coefficient for the natural log of depth (lnDEPTH). The coefficient is positive (0.53) indicating that total well yield increases with total well depth, but incremental yield (yield per foot of well depth) decreases with well depth. This result is in accordance with the physical relation, where water-bearing fractures generally decrease in size and number with depth (U.S. Geological Survey, 1984, p. 304). The demand-driven coefficient for lnDEPTH (-0.88) was negative and could not possibly represent the physical relation.

The coefficient of slope of the land surface (variable SLOPE, a unitless percent) is negative (-0.014), which indicates steep slopes tend to have low yields. Similarly, the land-surface elevation at each well (ELEV24, in feet above sea level) is derived from 1:24,000- or 1:25,000-scale maps with a negative coefficient of -0.00012, indicating that high elevations tend to have low yields. The curvature of the land surface determined from the 1:250,000-scale regional DEM (CURV250) has a negative coefficient of -2.15, indicating that, at a regional scale, concave downward surfaces are associated with low yields and concave upward surfaces are associated with higher-than-average yields.

Distance to the nearest waterbody, determined using the 1:24,000-scale Digital Line Graph (DLG), was negatively related to lnYIELD, with a coefficient of -0.00012. This relation indicates that wells further away from waterbodies tend to have lower than average yields, although other variables in the regression equation must be taken into consideration for estimating yields. The drainage area to the well, expressed as the natural log of the number of 1:24,000-scale DEM cells upgradient of the well (including the well cell, lnACCUM), has a positive coefficient of 0.024. This result indicates that large upgradient drainage areas are associated with higher-than-average yields, although other variables in the equation must be taken into consideration.

Table 2. Regression equation used by this study

[Geologic units from Lyons and others (1997), Walsh and Clark (1999), and Thomas R. Armstrong and William C. Burton (U.S. Geological Survey, written commun., 2000); *, dichotomous indicator variable; DEM, Digital elevation model; DLG, Digital line graph]

	Coefficient	Variable	Explanation of variable name model mean square error = 1.84
		lnYIELD	Dependent variable the natural log of yield, in gallons per minute
=	0.53	lnDEPTH	Natural log of well depth, in feet
			Topographic variables including proximity to surface water
–	.014	SLOPE	Percent slope from 1:24,000-scale DEM
–	.00012	ELEV24	Elevation derived from 1:24,000-scale map, in feet above sea level
–	2.15	CURV250	Curvature from 1:250,000-scale DEM
–	.00012	DISTH2O	Distance to water, in feet, determined from the 1:24,000-scale DLG
+	.024	lnACCUM	Natural log of the number of 1:24,000-scale DEM grid cells uphill of the well
+	.81	L7DRAW3	*Valley bottoms (1:24,000-scale DEM) lithologic group 7 (Massabesic Gneiss Complex and Breakfast Hill Granite)
			Major lithologic group
–	.27	LITHO2	*Major lithologic group 2, well-foliated igneous rocks
			Lineaments
+	1.23	D180BUF	*Domain fracture-correlated lineaments from high-altitude 1:80,000-scale photographs
+	.37	ALFRACP	*Discrete fracture-correlated lineaments in plutons from four platforms
+	.44	PAC15	*Statewide lineaments in plutons from angle category 15 (N30 ^{o}W to N40 ^{o}W)
			Geologic units from the Pinardville and Windham quadrangles
–	.14	SOB	*Berwick Formation
+	.61	SOBW	*Well-bedded Berwick Formation
+	.36	ZMLG	*Layered paragneiss and orthogneiss
+	1.55	ZMA	*Amphibolite gneiss
+	.86	PZMG	*Layered Migmatite
–	.44	PDG	*Damon Pond Granite
+	.41	DSG	*Spaulding Tonalite
+	.44	AXMBUF	*Within 2,000 ft of the axis of the Massabesic Gneiss Complex anticline
			Geologic units from the State map
+	.18	RKTYP101	*Perry Mountain Formation (Sp)
–	1.32	RKTYP102	*Undivided Perry Mountain & Rangeley Formations (Spr)
+	.36	RKTYP103	*Eliot Formation (SOe)
–	1.14	RKTYP105	*Perry Mountain Formation member (Spvx)
+	.62	RKTYP107	*Kittery Formation (SOk)
–	1.13	RKTYP112	*Rangeley Formation member (Srup)
–	.36	RKTYP120	*Massabesic Gneiss Complex (Zmz)
–	.20	RKTYP121	*Member of Berwick Formation (SObc)
–	.52	RKTYP122	*Calef member of Eliot Formation (SOec)
–	.37	RKTYP125	*Ammonoosic Volcanics member (Oal)
+	0.48	RKTYP139	*Granite, granodiorite, trondhjemite (Oo1-3A)
+	1.66	RKTYP174	*Rye Complex (OZrz)
+	.34	RKTYP45	*Two Mica Granite (D1m)
+	1.96	RKTYP53	*Metamorphosed gabbro, diorite, and basalt dikes (DS9)
+	.22	RKTYP57	*Concord Granite (Dc1m)

Table 2. Regression equation used by this study—Continued

	Coefficient	Variable	Explanation of variable name model mean square error = 1.84
			Geologic units from the State map--Cont.
-	1.28	RKTYP58	*Gile Mountain Formation (Dg)
-	.15	RKTYP66	*Undivided Littleton Formation (Dl)
-	1.77	RKTYP72	*Littleton Formation member with volcanic lentils (Dlv)
-	.30	RKTYP83	*Winnipesaukee Tonalite (DW3A)
+	.66	RKTYP90	*Undivided Frontenac Formation undivided (Sfr)
+	.89	RKTYP91	*Frontenac Formation member (Sfrb)
+	.61	RKTYP92	*Frontenac Formation member (Srfc)
			Well characteristics
+	.49	CABLE	*Well drilled with a cable tool rig
+	.38	COMERCL	*Commercial supply well
+	.93	PUBLIC	*Public supply well (note: domestic wells are those not identified as either public or commercial)
–	.80		End of equation

Swales or valley bottoms (L7DRAW3) were identified using the 1:24,000-scale DEM. In major lithologic unit 7 (LITHO7) (Massabesic Gneiss Complex and Breakfast Hill Granite) a positive coefficient of 0.81 indicates that these sites tend to have higher-than-average well yields.

The variable identifying well sites within 100 ft of lineaments that were identified using the high-altitude (1:80,000-scale) aerial photography, and also correlated with the primary fracture direction (regional analysis) (D180BUF), had a large positive coefficient of 1.23. This result indicates that wells near these lineaments tend to have high well yields.

Similar to the description above, the indicator variable identifying well sites in plutons (lithologic groups 1, 2, and 7) within 100 ft of lineaments (ALFRACP), which were correlated with observed fractures in outcrops, measured within 1,000 ft of these lineaments, and identified by use of (1) Landsat; (2) high-altitude aerial photography; (3) color infra-red photography; or (4) topographic maps, had a positive coefficient of 0.37. This result indicates an association of these lineaments to higher-than-average well yields.

The variable identifying well sites in plutons within 100 ft of lineaments trending N35^{o}W ±5 degrees (PAC15) that were identified using Landsat, or high-altitude or low-altitude aerial photography, had a positive coefficient of 0.44. PAC15 sites are associated with high yields. Statewide, lineaments in this orientation seem to represent a group of fractures with higher-than-average yields, possibly the result of openness of these fractures or a similar method of formation. The degree of openness and hydrologic connectivity to horizontal fractures may be greater in the plutons than in metasedimentary rocks.

Seven detailed geologic map units (Zma, SOb, SObw, Zmlg, PZmg, Pdg, and Dsg), identified during the detailed geologic mapping of the Pinardville and Windham quadrangles, were significant in the regression model. Well yields were either higher or lower than average depending on the geologic unit.

Major lithologic unit 2 (LITHO2), foliated plutons, had a negative coefficient of -0.27 indicating that these areas tend to have low yields. Twenty-two of the 154 geologic map units, depicted on the Bedrock Geologic Map of New Hampshire (Lyons and others, 1997), also were statistically significant at the 95-percent confidence level. These mapped units have either a positive or negative effect depending on the unit. For example, mapped units in major lithologic category 2, LITHO2 (fig. 3), also are significant in the regression model (RKTYP139, RKTYP174, RKTYP45, RKTYP53, RKTYP57, and RKTYP83, table 2). These represent local refinements to the generalized rock groupings. The effect of these units in LITHO2 is essentially to either add or subtract from the overall negative coefficient of major LITHO2 unit.

In addition to site characteristics, certain well characteristics also were determined to be statistically significant. CABLE, which identifies whether or not a well was drilled using a cable tool rig, has a positive relation to well yield. High yields among these drilled wells may be the result of the drilling technique in which the bit is pounded into the earth, perhaps opening up or enhancing fractures. Also, there may be less clogging of fractures by bedrock cuttings when the well is drilled with a cable-tool rig.

Wells drilled for commercial or public water supply (COMERCL and PUBLIC) also are associated with high yields. These yields may be related to well construction characteristics, such as well diameter, that may increase the well yield at a given site to meet the demand. The fact that these wells tend to be deeper than private wells already is accounted for in the first stage of the regression.

Other major variables, which individually were significant but dropped out as new variables were introduced, include (1) the distance of a well to any lineament in a pluton, (2) thickness of overburden, and (3) whether a well penetrates through stratified drift. With the introduction of variables, such as distance to surface-water bodies (DISTH2O) and upgradient drainage area (lnACCUM), these three variables dropped out of the model because of shared variance. Variables DISTH2O and lnACCUM were statistically more significant than those values that were dropped. The problem of shared variance is one of exclusion. If two variables share variance and both are included in the model and both remain significant, then there is no modeling problem. If, however, variables are excluded from the model because they fail to meet the 95-percent confidence level because of shared variance, then there is the potential to exclude variables that should be included. This is an unavoidable consequence of any regression analysis. The model is stochastic or statistical, not a process-oriented geologic, and cannot account for all physical processes involved, which may be best explained by some excluded variable.

Regression Model Testing

The statewide analysis of bedrock yields by the USGS involved developing a complex regression model. Large data sets were used to examine geospatial data. The resultant model then was subjected to six statistical tests (appendix C), and the results are summarized in the following paragraphs:

1. First, a test was conducted to determine if the use of instrumental variables was necessary. A test was applied to evaluate the use of the instrumental-variable technique to eliminate the effect of the inherent relation between well yield and depth. In applying this test, lnDEPTH was endogenous and a two-stage regression model using instrumental variables was appropriate.
2. Second, the appropriateness of the form of the model was tested. A test of the model specification revealed that nine variables are heteroscedastic. That is, the distribution of the model residuals varies depending on the value of these nine variables. The effects of heteroscedastic variables on model probabilities is discussed in appendix C. The form of the model was appropriate.
3. A test for normality of the residuals indicated that the graphical distribution is nearly normally distributed; however, at the 95-percent confidence level, the residuals are not normally distributed. For practical purposes, the model was considered appropriate for estimating probabilities of equaling or exceeding a given yield.
4. Declustering of the data (statistically deleting some of the excess data where there is an over abundance) was not necessary on the basis of a spatial test of the residuals for the Pinardville quadrangle. Pinardville is the quadrangle with the highest density of well data in the State. Results of the spatial test did not show spatial significance, which means that the data are not related to those of neighboring wells in the most clustered area of the map, and would not be a potential source of bias.
5. A verification test of model results was done. A summary Chow test (Chow, 1960) was used to compare the verification and primary data sets. The results indicate that the data sets are statistically distinguishable from one another by the model; they do not behave identically. However, a single variable, PZmg, accounts for most of the difference between the verification data set and the primary data-set (with coefficients of opposite sign between the two data sets).
6. A sixth test, unique to the use of instrumental variables, was a test of finite sample bias. The test results showed no indication of significant finite-sample bias.

Alternative models expressing yield and depth were considered. The final form of the selected model was considered appropriate because of the following reasons:

1. The model adjusts for the severe endogenous bias associated with depth.
2. The residuals are, for all practical purposes, log normally distributed indicating that the decision to express yield and depth in log space was appropriate.
3. The errors are multiplicative (higher errors at higher yields) further indicating that a model with lnYIELD as the dependent variable (as opposed to yield) was appropriate.
4. The specification tests demonstrate that the functional form was appropriate for the continuous variables.
5. The functional form of the model was not an issue because only dichotomous variables were identified with the specification test. Heteroscedasticity was found for some dichotomous variables, but this can be considered when interpreting the results.

Yield-Probability Forecasting

The probability of obtaining a given yield can be forecast statewide for a given well depth. A statewide GRID of probabilities of equaling or exceeding 40 gal/min when a well depth of 400 ft is drilled is presented in figure 9. A well depth of 400 ft was used because this has become a common depth at which a driller stops drilling if the desired yield has not been reached.

Maps of Yield Probability

To apply the results of the regression model, a map is needed. Maps (figs. 9, 10, and 11) clearly show areas of various yield probabilities. Users then can use probability criteria to select areas where other techniques, such as geophysical techniques, can be applied to further refine site selection for drilling in the bedrock aquifer. Results of tests of geophysical techniques for ground-water exploration in the fractured-bedrock aquifer of New Hampshire are presented in a companion report by Degnan and others (2001).

The regression model developed during this study was used to produce yield-probability maps, statewide (fig. 9) and for the Pinardville and Windham quadrangles, N.H. (plates 1 and 2). The maps were created by the following procedures. A grid of points, with a 98-ft (30-m) spacing, was created. For each point, all predictive parameters were compiled and model results applied. For each point, a probability of equaling or exceeding a given yield was calculated. The grid of points then was converted to a raster GRID (using ARC/INFO GRID software) and plotted on a map. The probability at each GRID cell was determined by using equations 1 and 2 and the predicted value of lnYIELD that was computed by the regression equation. The procedure used to estimate cell probability, expressed as a percentage, is

$$\text{Cell Probability} = 100 * (1 - PROBNORM\ (Z)), \quad (1)$$

where

PROBNORM (SAS computer software, 1996) is a function that returns the probability that an observation from a standard normal distribution is less than *Z*,

Z is a function that defines the point where the conditional yield (to be equaled or exceeded) is relative to the predicted distributions of yields for each cell, and

Z is normalized by the standard deviation of predicted yields for each cell.

$$Z = (\ln(C)) - \hat{Y} / \sqrt{(MSE + BETAVAR)}, \quad (2)$$

where

C is the conditional yield to be equaled or exceeded,

$\hat{Y}$ is the predicted natural log of yield for the cell (lnYIELD),

MSE is the model mean square error, and

BETAVAR is the variance associated with the coefficients. (This involves the covariance matrix and is dependent on the magnitude of the predictor variable).

The map data are displayed in shades of color on the basis of the value of probability (figs. 9, 10A, and 11A, and plates 1 and 2). The effect of topography is reflected in the probability zones. Low probabilities

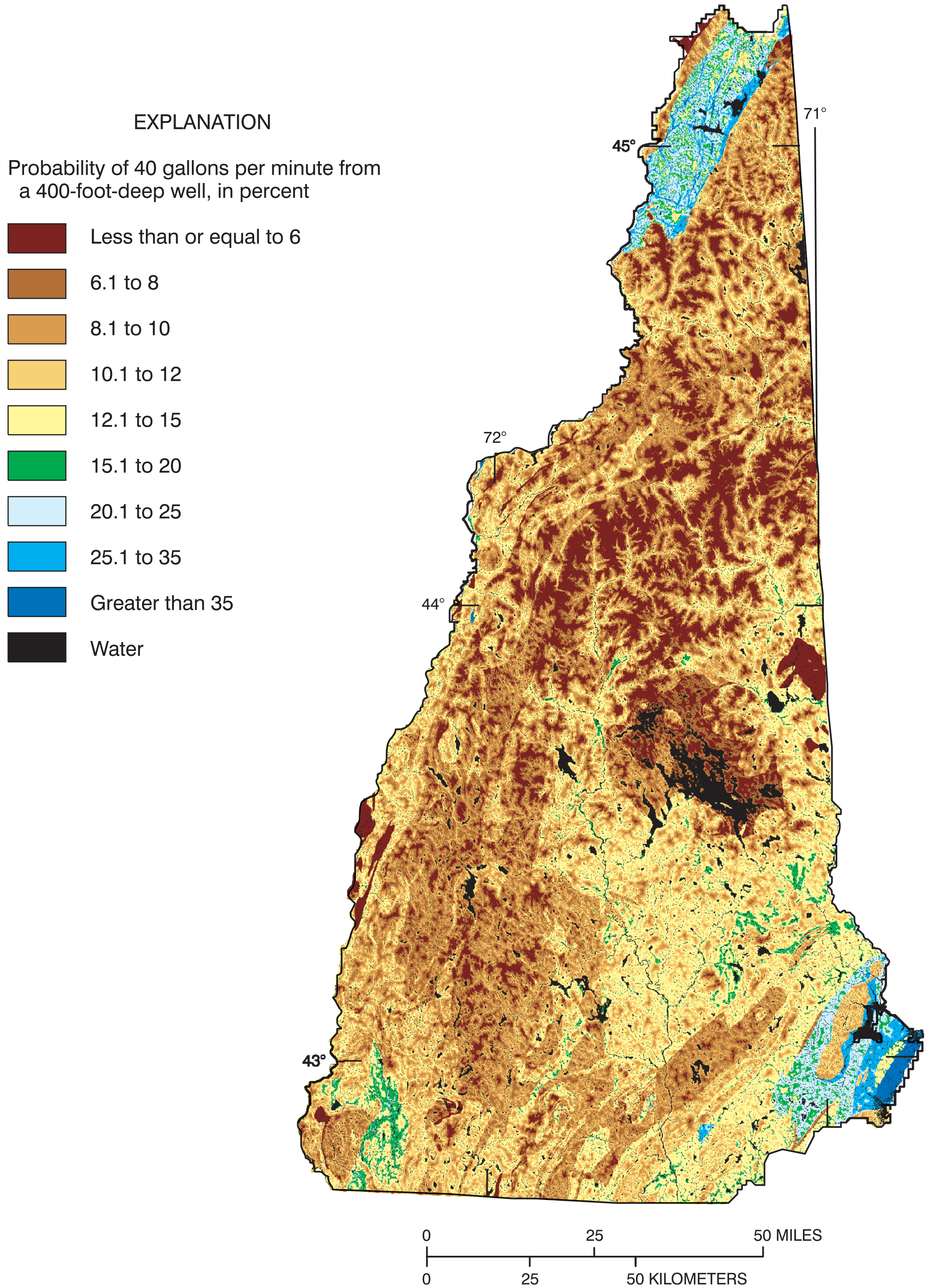

Figure 9. Model results of the probability of obtaining 40 gallons of water per minute from a 400-foot-deep well in New Hampshire.

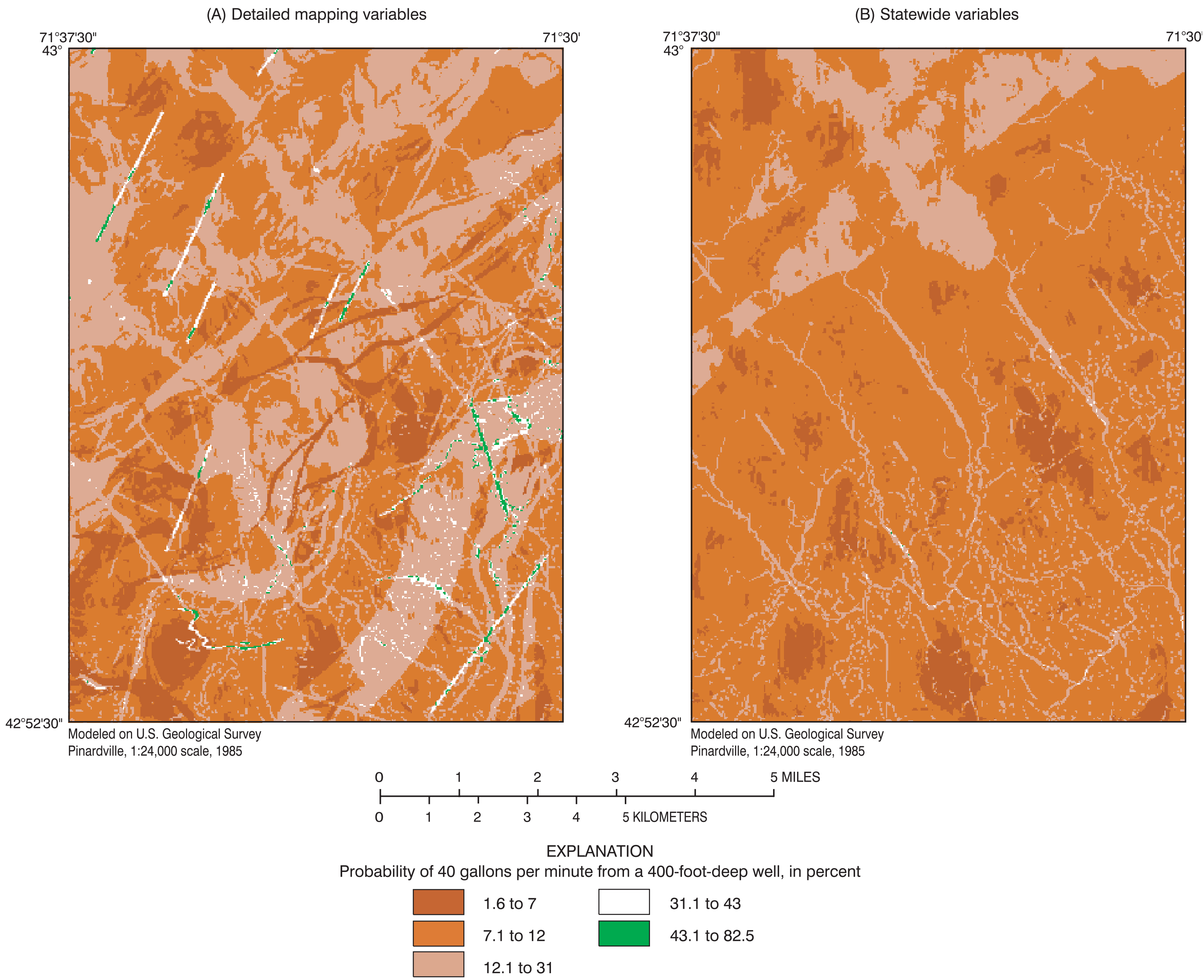

Figure 10. Model results of the probability of obtaining 40 gallons of water per minute or more from a 400-foot-deep well for the Pinardville quadrangle, New Hampshire, with (A) detailed mapping variables included and (B) only the statewide variables included

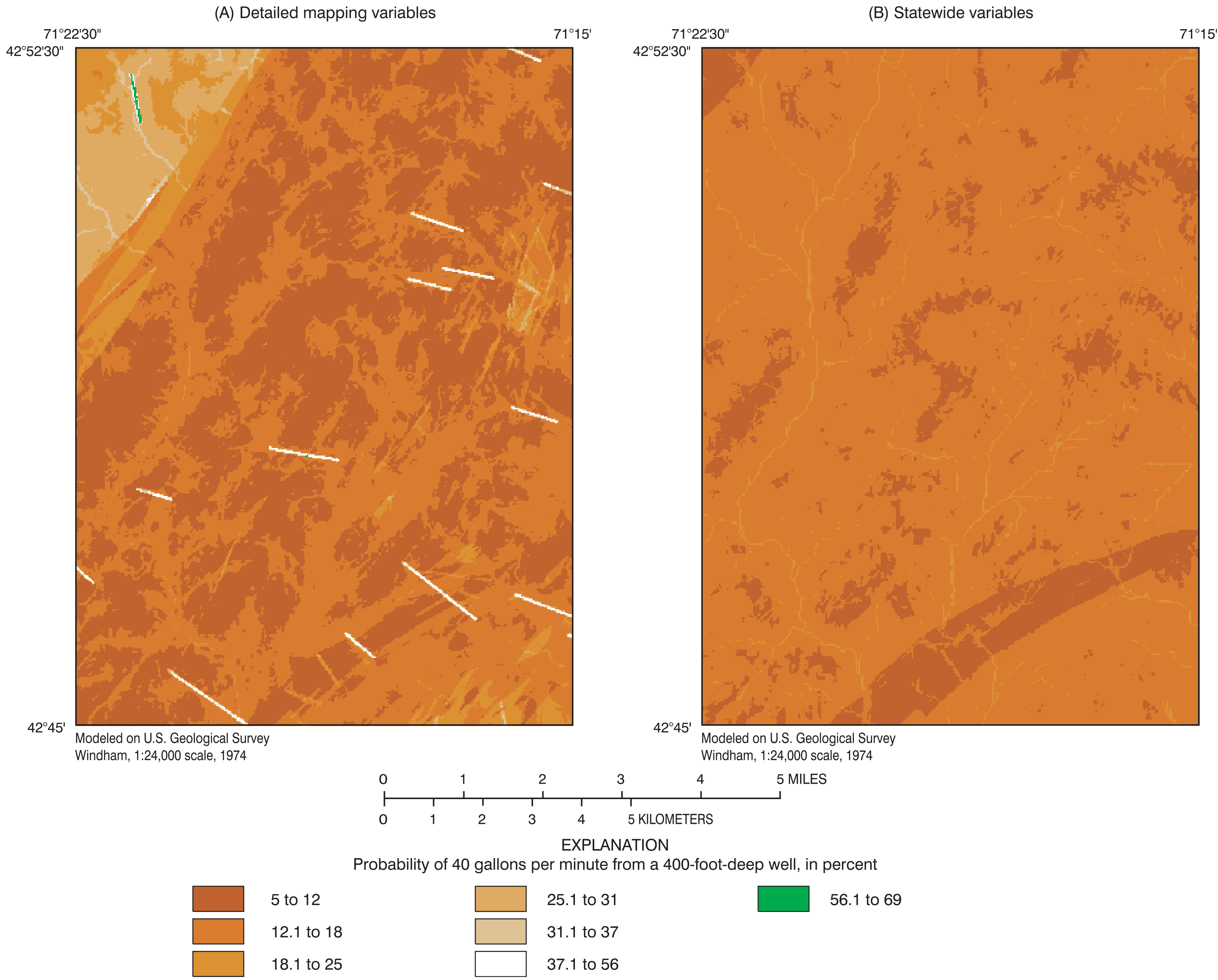

Figure 11. Model results of the probability of obtaining 40 gallons of water per minute or more from a 400-foot-deep well for the Windham quadrangle, New Hampshire, with (A) detailed mapping variables included and (B) only the statewide variables included.

on the steep slopes are represented by dark shades of orange. Lineaments and topographic swales (determined from the 1:24,000-scale DEM) are clearly highlighted in white or green. This result is especially true of the lineaments identified using the high-altitude aerial photography that are correlated with the primary fracture direction. These lineaments define most of the locations with the highest probability areas, which indicates that the regression model is sensitive to photolinear variables despite measurement error and shared variance. Results of this analysis indicate that using a filtered subset of the lineaments (based on lineament orientation and correlation with nearby fracture orientations) to identify potential high-probability areas is feasible elsewhere in New Hampshire.

A broad swath of high probabilities associated with the anticlinal axis of the Massabesic Gneiss Complex (figs. 4 and 10A), across the southeastern quadrant of the map, is prominent. Likewise, the positive (light colors) and negative effect (dark orange) of the various detailed mapped units identified in the model are clearly evident. For example, the bedrock geologic units of the amphibolite gneiss (Zma), layered migmatite (PZmg), and Spaulding Tonalite (Dsg) (fig. 4) all are highlighted as zones of high-yield probabilities (fig. 10A and plate 1).

Comparison of Regional Variables with Statewide Variables

To measure the added value of field-based methods for well-yield-probability forecasting, the regression model was run with and without the variables developed from the detailed geologic mapping and fracture-correlated lineaments (Zma, SOb, SObw, Zmlg, PZmg, Pdg, Dsg, AXMBUF, D180BUF, AND ALFRACP (table 2)). Four probability maps were produced (figs. 10 and 11) for the Pinardville and Windham quadrangles, one with and one without the added variables for each quadrangle. These maps show the benefit of conducting additional detailed geologic mapping.

The dark zones in the northwestern quadrant of plate 1 and figures 10A and 10B are associated with the steep slopes and concave downward curvature of the Uncanoonuc Mountains. Likewise, valleys and swales are evident as light zones in the Massabesic Gneissic Complex in the southeastern half of the plate. Lineaments trending N35^{o}W in the igneous rocks (plutons) also are shown as distinct lines. Fracture-correlated lineaments (trending in various directions), the zone of horizontal jointing associated with the anticlinal axis of the Massabesic Gneiss Complex, and the effects of five detailed geologic units only are evident on the map that included the detailed mapping variables (fig. 10A).

Well-yield probability maps with and without the detailed mapping variables can be compared quantitatively (table 3). At any selected probability value, more sites are identified with the inclusion of the detailed mapping. For example, with a 20-percent probability of obtaining 40 gal/min from a 400-ft-deep well, more than 3 times as many sites could be located by including the variables from the detailed mapping. As the criterion is restricted to higher percentage probabilities, the benefit of using detailed mapping to identify areas increases. For example, at and above 45-percent probability, the model with the excluded detailed geologic mapping variables does not identify any sites, whereas the model with the variables included continues to identify sites (table 3). The amount of detail in the probability maps also can be summarized in terms of the variance (or standard deviation) of predicted probability values. For the Pinardville quadrangle, the variance increased from 12.43 to 40.56 with the inclusion of the detailed variables and the standard deviation increased from 3.53 to 6.37. These wide ranges in estimated probabilities indicate that more high-probability sites are identified by including the detailed mapping variables.

A similar comparison, as described in the previous paragraph, can be made in the Windham quadrangle (fig. 11). The effects of the statewide topographic variables are evident in figures 11A and 11B. The dark orange zones with highly irregular shapes on the maps reflect the negative effects of hills. A dark orange zone also indicates the low probability associated with the Ayer granodiorite, part of LITHO2 (fig. 3), along the southern boundary of the quadrangle (figs. 11A and B, plate 2) near the center. Similarly, the light zones associated with RKTYPE45 (D1m, the Two Mica Granite of northern and southeastern New Hampshire) are evident on figures 5 (unit Dg) and 11A. Proximity to water is shown as light shades of color on both maps. The well-bedded subunit of the Berwick Formation (SObw) has high probabilities (lighter-shaded zone of the northwest corner (fig. 11A), which is not identified in the statewide map (fig. 11B).

Table 3. Percentage of area identified as having a probability of obtaining 40 gallons per minute or more from a 400-foot-well depth with detailed geologic mapping variables included and excluded for the Pinardville quadrangle

Probability of obtaining 40 gallons per minute or more with 400-foot-well depth, in percent	Percentage of quadrangle meeting probability criteria with detailed geologic mapping variables	
	Included	Excluded
Pinardville quadrangle, New Hampshire		
20	7.94	2.33
25	3.93	.80
30	2.22	.10
35	1.33	.02
40	0.72	.01
45	.39	0
50	.21	0
55	.11	0
60	.06	0
65	.02	0
70	.013	0
75	.004	0
80	.002	0

Various lineaments associated with the igneous rock show up as linear patterns of increased probability on both quadrangles (figs. 10A and 11A, plates 1 and 2). The highest probabilities identified in linear pattern on figures 10A and 11A, and plates 1 and 2 are from lineaments identified using the high-altitude aerial photography; these lineaments are correlated with the primary fracture direction (regardless of bedrock type).

A quantitative comparison of probability maps for the Windham quadrangle is given in table 4. More sites are identified at selected probability criteria with the detailed mapping. For this quadrangle, however, the benefits of the detailed geologic mapping seems to be even greater than for the Pinardville quadrangle. For example, 23 times as many sites are located with a 20-percent probability of obtaining 40 gal/min with a 400-ft-deep well, when the detailed geologic mapping variables were included. At or above 25-percent probability, the model, that excludes variables, fails to identify any sites. The model with the variables included continues to identify sites (table 4). The amount of detail in the probability maps also can be summarized in terms of the variance (or standard deviation) among predicted probability values. For the Windham quadrangle, the variance increased from 3.64 to 26.74, with the inclusion of the detailed variables, and the standard deviation increased from 1.91 to 5.17.

Table 4. Percentage of area identified as having a probability of obtaining 40 gallons per minute or more from a 400-foot-well depth with detailed geologic mapping variables included and excluded for the Windham quadrangle

Probability of obtaining 40 gallons per minute or more with 400-foot-well depth, in percent	Percentage of quadrangle meeting probability criteria with detailed geologic mapping variables	
	Included	Excluded
Windham quadrangle, New Hampshire		
20	10.23	0.44
25	6.28	0
30	1.33	0
35	.62	0
40	.40	0
45	.18	0
50	.065	0

SUMMARY AND CONCLUSIONS

The New Hampshire Bedrock Aquifer Assessment, done in cooperation with the New Hampshire Department of Environmental Services, was designed to provide information that can be used by communities, industry, professional consultants, and other interests to evaluate the ground-water development potential of the fractured-bedrock aquifer. This assessment was done at three scales—statewide, regional, and local—to uncover relations that potentially will increase the probability of successfully locating high-yield water supplies in the fractured-bedrock aquifer throughout New Hampshire. The statewide scale was designed as a reconnaissance-level investigation of bedrock well yield in relation to bedrock type, lineament characteristics, topography, and other well-site and construction characteristics. The regional scale adds field-based geologic-data collection, additional lineament data, and lineament filtering to assess what effect quadrangle-scale data acquisition has on the resulting well-yield relations. Local investigations were done to assess the effectiveness of geophysical tools for identifying the location and orientation of fracture zones. Results of the statewide- and regional-scale investigations are described in this report. Results of the local investigations are presented in a companion report.

In investigating the factors that relate to high-yield bedrock-aquifer sites, many possible variables were examined. Statewide data sets were compiled from available sources or were created for this study. The primary information that was analyzed included well information, bedrock lithology, surficial geology, lineaments, topography, and various derivatives of information in those data sets.

Additionally, more detailed geologic, fracture, and lineament data were collected for the Pinardville and Windham, New Hampshire quadrangles. These quadrangles were selected because they represent differing geohydrologic settings and have the largest number of georeferenced bedrock wells in the State. Geologic mapping at the 1:24,000 scale provided an accurate distribution of lithologies and data on brittle fracture and ductile-structure orientation of bedrock for correlation with lineaments. The purpose of this regional-scale investigation was to determine the degree to which predictive well-yield relations, developed as part of the statewide reconnaissance investigation, can be improved with quadrangle-scale mapping.

Detailed information on water wells constructed in the State since 1984 is collected and maintained in a computer database by NHDES. Many of these wells have been field-inventoried to obtain accurate locations. Basic well data are reported by drillers and include the following six characteristics used in this study: well yield, well depth, well use, method of construction, date drilled, and depth to bedrock (or length of casing). The well data set compiled by NHDES contains complete records, including geographic coordinates for 20,308 of the total number of wells (over 70,000) reported since 1984. A randomly selected subset (20 percent or 4,050 wells) of the available well data were reserved for model evaluation. The availability of this large data set provided an opportunity for statistical analysis of variables related to bedrock well yields. Well yields in the database ranged from zero to greater than 500 gal/min.

Multivariate regression was selected as the primary method of analysis because it is the most efficient statistical tool for predicting a single variable (such as the natural log of well yield) with many potentially independent variables. The dependent variable that was analyzed in this study was the natural log of the reported well yield. Well yield, however, is in part a function of demand for water by the well owner. This restriction introduces a major problem in using well yield as the dependent variable in a multivariate-regression analysis. Generally, a well only is drilled as deep as is needed to meet a specific yield. Thus, the reported yield is not necessarily the maximum potential yield for that site. Because well yield is partially a function of demand, an innovative technique that involves the use of instrumental variables was used. Instrumental variables that are correlated to demand-driven yield but uncorrelated with random variation in the physical yield-depth relation are (1) year drilled, (2) well driller, and (3) median household income based on the 1990 census.

Results of the regression model show the following:

1. The effect of the instrumental variable approach is apparent in the calculated regression coefficient for the natural log of depth (lnDEPTH). The regression coefficient is positive (0.53) indicating that total well yield increases with total well depth, but incremental yield (yield per foot of well depth) decreases with well depth.
   The previously demand-driven negative coefficient (-0.88) was corrected by a coefficient that represents the physical relation. Presumably at some depth (hundreds to thousands of feet) this relation ceases because of lithostatic pressure.
2. Slope of the land surface determined from the 1:24,000 scale DEM, has a negative coefficient implying that yields tend to decrease as slopes increase. Similarly, the elevation of the land surface at each well, derived from 1:24,000 or 1:25,000-scale maps, has a negative coefficient of -0.00012, indicating that yields tend to decrease as elevation (in feet above sea level) increases. The curvature of the land surface determined from the 1:250,000-scale regional DEM has a negative coefficient of –2.15, indicating that, at a regional scale, wells on hilltops tend to have low yields and wells in valleys tend to have higher-than-average yields. The data indicate that valleys have more saturated fractures than hilltops.
3. The distance to the nearest waterbody, determined using the 1:24,000 scale Digital Line Graph data (DLG), is negatively related to lnYIELD. This means that wells farther away from waterbodies tend to have lower yields than those closest to waterbodies.
4. The drainage area to the well, expressed as the natural log of the number of 1:24,000-scale DEM cells upgradient of the well (including the well

cell) (lnACCUM) has a positive coefficient of 0.024. This value indicates that wells with large upgradient drainage areas tend to have higher-than-average yields.

5. Swales or valley bottoms, identified using the 1:24,000 scale DEM, in the Massabesic Gneiss Complex and Breakfast Hill Granite, tend to have higher-than-average yields.
6. Major lithologic unit 2, foliated plutons, tended to have lower than average yields.
7. Sites within 100 ft of lineaments, identified using the high altitude (1:80,000-scale) aerial photography that are correlated with the primary fracture direction (regional analysis), have a large positive coefficient of 1.23. This value indicates a strong association between these lineaments and higher-than-average well yields.
8. The quadrangle-scale well sites in plutons within 100 ft of fracture-correlated lineaments (discrete analysis), defined as a subset of all lineaments identified using Landsat, high-altitude aerial photography, color infra-red photography, or topographic maps, have higher-than-average yields. Perhaps sheeting (near-horizontal) fractures in plutons create more hydrologic connections between wells and adjacent lineaments.
9. Sites in plutons within 100 ft of lineaments trending N35^{o}W ±5 degrees, identified using Landsat, or high- or low-altitude aerial photography, have a positive coefficient of 0.44, indicating higher-than-average yields.
10. Seven geologic map units (Zma, SOb, SObw, Zmlg, PZmg, Pdg, and Dsg) identified by detailed geologic mapping of the Pinardville and Windham quadrangles, had a relation to well yield. Well yields were statistically above or below average depending on the geologic unit.
11. Twenty-two geologic map units on the Bedrock Geologic Map of New Hampshire also had a relation to well yield. These had either a positive or negative relation to yield depending on the type of unit.
12. In addition to the above site characteristics, certain well-construction characteristics also were determined to be statistically significant. Wells drilled using a cable-tool rig were positively related to well yield.
13. Wells drilled for commercial or public-supply purposes also tended to have higher-than-average yields.

A statistical relation predicting well yield was developed to produce yield-probability maps. Plates show the yield probabilities for the Pinardville and Windham quadrangles. Probabilities are color coded on these maps. **The results of the yield-probability maps** indicate the following:

1. The effect of topography is reflected in the probabilities. Probabilities of high yields decrease on the steep slopes.
2. Specific categories of lineaments result in linear patterns of high probabilities on the maps. This is especially true of the lineaments identified using the high-altitude aerial photography (1:80,000 scale) that are correlated with the primary fracture direction. These lineaments define most of the locations with the highest probabilities in the Pinardville and Windham quadrangles. As a result of this analysis, the use of a filtered subset of lineaments appears to be a promising technique for identifying high-probability areas elsewhere in New Hampshire.
3. A broad swath of high probabilities is associated with the anticlinal axis of the Massabesic Gneiss Complex across the southeastern quadrant of the Pinardville quadrangle. Likewise, the positive and negative effect of the variables representing various detailed lithologies, identified in the model, are evident. For example, the bedrock geologic units of the amphibolite gneiss (Zma), layered migmatite (PZmg), and Spaulding Tonalite (Dsg) are all represented as zones of high-yield probabilities.
4. To measure the added value of using field-based methods for well-yield-probability forecasting, probabilities were estimated with and without the variables developed from the detailed geologic mapping and fracture-correlated lineaments. The probability maps provide a visual comparison for evaluating the benefit of including additional detailed geologic mapping variables. The results of including these variables for each quadrangle are as follows:

4A. In the Pinardville quadrangle, more sites are identified at selected probability criteria with detailed mapping variables. For example, using the information from the detailed mapping, more than 3 times as many sites are associated with a 20-percent probability of obtaining 40 gal/min from a 400-ft-deep well. As probability criteria are restricted to percentages higher than 20 percent, the advantage of the detailed mapping in identifying additional areas of high-yield increases. At and above 45-percent probability, the model using just the statewide variables fails to identify any sites, whereas the model using the detailed mapping variables continues to identify sites.

4B. Similar results were found for the Windham quadrangle where more sites were identified using the detailed mapping variables. For this quadrangle, however, the increased percentage of area with a high probability of obtaining 40 gal/min or more with detailed geologic mapping variables included, is even greater than for the Pinardville quadrangle. In areas with a 20-percent or more probability of obtaining 40 gal/min from a 400-ft-deep well, roughly 23 times as many sites are identified using the information from the detailed mapping. At or above 25-percent probability, the model with just the statewide variables does not identify any sites, whereas the model with the detailed mapping variables continues to identify sites.

The amount of detail in the probability maps also can be expressed in terms of the variance (or standard deviation) among predicted probabilities. For the Pinardville quadrangle, the variance describing the range of predicted probability values increased with the inclusion of the detailed mapping variables, from 12.43 to 40.56, and the standard deviation increased from 3.53 to 6.37. For the Windham quadrangle, the variance increased from 3.64 to 26.74, and the standard deviation increased from 1.91 to 5.17. These wide ranges in probabilities indicate that more high-probability sites can be identified by including the detailed mapping. Results from both quadrangles clearly demonstrate the advantage of using detailed geologic mapping to identify potential new ground-water supplies in the fractured-bedrock aquifer of New Hampshire.

SELECTED REFERENCES

Bates, R.L., and Jackson, J.A., 1980, Glossary of geology: Falls Church, Va., American Geological Institute, 751 p.

Blanchet, P.H., 1957, Development of fracture analysis as exploration method: Bulletin of the American Association of Petroleum Geologists, v. 41, no. 8, p. 1748-1759.

Boudette, E.L., 1990, The geology of New Hampshire: Rocks and Minerals, v. 65, p. 306-312.

Bound, John, Jaeger, D.A., and Baker, R.M., 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak: Journal of the American Statistical Association, v. 90, no. 430, p. 443-50.

Brown, C.W., 1961, Comparison of joints, faults, and airphoto linears: American Association of Petroleum Geologists Bulletin, v. 45, p. 1888-1892.

Caswell, Bradford, 1987a, Results of building a 12-inch-diameter production well in granite: Water Well Journal, March 1987, p. 38-40.

———1987b, Anisotropy in fractured bedrock aquifers: It can be helpful: Water Well Journal, September 1987, p. 32-33.

Chormann, F.H., Jr., 1990, Bedrock water wells in New Hampshire, a statistical summary of the 1984-1990 inventory: New Hampshire Department of Environmental Services, NHDES-WRD-90-3, 52 p.

Chormann, F.H., Jr., and Moore, R.B., 1998, The New Hampshire bedrock aquifer resource assessment: The latest generation of aquifer mapping: New Hampshire Water Works Association, Summer Journal 1998, p. 13-25.

Chow, G.C., 1960, Tests of equality between sets of coefficients in two linear regressions: Econometrica, v. 28, p. 591-605.

Clark, S.F., Jr., Ferguson, E.W., Picard, M.Z., and Moore, R.B., 1996, Lineament map of area 2 of the New Hampshire bedrock aquifer assessment, south-central New Hampshire: U.S. Geological Survey Open-File Report 96-490, scale:1:48,000.

———1998, Lineament map of area 4 of the New Hampshire bedrock aquifer assessment, southeast-central New Hampshire: U.S. Geological Survey Open-File Report 97-763, scale 1:48,000.

———1998, Lineament map of area 6 of the New Hampshire bedrock aquifer assessment, central New Hampshire: U.S. Geological Survey Open-File Report 98-189, scale 1:48,000.

Clark, S.F., Jr., Ferguson, E.W., Short, H.A., Marcoux, G.J., and Moore, R.B., 1999, Lineament map of area 9 of the New Hampshire bedrock aquifer assessment, north-central New Hampshire: U.S. Geological Survey Open-File Report 99-63, scale 1:48,000.

Clark, S.F., Jr., Moore, R.B., Ferguson, E.W., and Picard, M.Z., 1996, Criteria and methods for fracture-trace analysis of the New Hampshire bedrock aquifer: U.S. Geological Survey Open-File Report 96-479, 12 p.

Clark, S.F., Jr., Short, H.A., Ferguson, E.W., and Moore, R.B., 1998, Lineament map of area 7 of the New Hampshire bedrock aquifer assessment, west-central New Hampshire: U.S. Geological Survey Open-File Report 98-190, scale 1:48,000.

Cressie, Noel, A.C., 1991, Statistics for spatial data: New York, John Wiley & Sons, Inc., 900 p.

Daniel, C.C., III, 1989, Statistical analysis relating well yield to construction practices and siting of wells in the Piedmont and Blue Ridge provinces of North Carolina: U.S. Geological Survey Water-Supply Paper 2341-A, 27 p.

———1990, Evaluation of site-selection criteria, well design, monitoring techniques, and cost analysis for a ground-water supply in Piedmont crystalline rocks, North Carolina: U.S. Geological Survey Water-Supply Paper 234-B, 35 p.

Daniel, C.C., III, Sharpless, N.B., 1983, Ground-water supply potential and procedures for well-site selection, Upper Cape Fear River Basin: North Carolina Department of Natural Resources and Community Development, 73 p.

Davidson, Russell, and MacKinnon, J.G., 1993, Estimation and inference in econometrics: New York, Oxford University Press, 874 p.

Degnan, J.R., and Moore, R.B., 2001, Identifying fracture zones in crystalline bedrock aquifers of New Hampshire using surface geophysical techniques *in* Conference on ground water data, Collection, Reliability, Access, and Manipulation of Basic Data, Nashville, Tenn., 2001, Proceedings: Nashville, Tenn., National Ground Water Association, p. 3-4.

Degnan, J.R., Moore, R.B., and Mack, T.J., 2001, Geophysical investigations of well fields to characterize fracture bedrock aquifers: U.S. Geological Survey Water-Resources Investigations Report 01-4183, 60 p.

Drew, L.J., Karlinger, M.R., Armstrong, T.R., and Moore, R.B., 1999, Relations between igneous and metamorphic rock, fracture patterns, and ground-water yield from the variography of water-well yields, Pinardville quadrangle, New Hampshire: Natural Resources Research, v. 8, no. 2, p. 137-152.

Dverstorp, Bjorn, and Anderson, Johan, 1989, Application of the discrete fracture network concept with field data: Possibilities of model calibration and validation: Water Resources Research, v. 25, p. 540-550.

Engelder, Terry, 1987, Joints and shear fractures in rock, *in* Atkinson, B.K., (ed.), Fracture Mechanics of Rock: London, Academic Press, Inc., p. 27-69.

Engelder, Terry, and Geiser, Peter, 1980, On the use of regional joint sets as trajectories of paleostress fields during the development of the Appalachian Plateau, New York: Journal of Geophysical Research, v. 85, no. B11, p. 6,319-6,341.

Ferguson, E.W., Clark, S.F., Jr., Marcoux, G.J., Short, H.A., and Moore, R.B., 1999, Lineament map of area 11 of the New Hampshire bedrock aquifer assessment, northeastern New Hampshire: U.S. Geological Survey Open-File Report 99-65, scale 1:48,000.

Ferguson, E.W., Clark, S.F., Jr., and Moore, R.B., 1997, Lineament map of area 1 of the New Hampshire bedrock aquifer assessment, southeastern New Hampshire: U.S. Geological Survey Open-File Report 96-489, scale 1:48,000.

Ferguson, E.W., Clark, S.F., Jr., Picard, M.Z., and Moore, R.B., 1998, Lineament map of area 3 of the New Hampshire bedrock aquifer assessment, eastern New Hampshire: U.S. Geological Survey Open-File Report 97-762, scale 1:48,000.

———1998, Lineament map of area 5 of the New Hampshire bedrock aquifer assessment, southwestern New Hampshire: U.S. Geological Survey Open-File Report 97-761, scale 1:48,000.

Ferguson, E.W., Clark, S.F., Jr., Short, H.A., Marcoux, G.J., and Moore, R.B., 1999, Lineament map of area 10 of the New Hampshire bedrock aquifer assessment, northwest-central New Hampshire: U.S. Geological Survey Open-File Report 99-64, scale 1:48,000.

Ferguson, E.W., Marcoux, G.J., Degnan, J.R., Clark, S.F., Jr., and Moore, R.B., 1999a, Lineament map of area 13 of the New Hampshire bedrock aquifer assessment, northern New Hampshire: U.S. Geological Survey Open-File Report 99-90, scale 1:48,000.

———1999b, Lineament map of area 14 of the New Hampshire bedrock aquifer assessment, northernmost New Hampshire: U.S. Geological Survey Open-File Report 99-91, scale 1:48,000.

Ferguson, E.W., Marcoux, G.J., Degnan, J.R., Clark, S.F., Jr., Short, H.A., and Moore, R.B., 1999, Lineament map of area 12 of the New Hampshire bedrock aquifer assessment, northwestern New Hampshire: U.S. Geological Survey Open-File Report 99-89, scale 1:48,000.

Ferguson, E.W., Short, H.A., Clark, S.F., Jr., Marcoux, G.J., and Moore, R.B., 1998, Lineament map of area 8 of the New Hampshire bedrock aquifer assessment, northeast-central New Hampshire: U.S. Geological Survey Open-File Report 98-191, scale 1:48,000.

Florquist, B.A., 1973, Techniques for locating water wells in fractured crystalline rocks: Groundwater, v. 11, no. 3, p. 26-28.

Gale, J.E., 1982, Assessing the permeability characteristics of fractured rock: Geological Society of America, Special Paper 189, p. 163-181.

Gates, W.C.B., 1997, The hydro-potential (HP) value: A rock classification technique for evaluation of the ground-water potential in fractured bedrock: Environmental & Engineering Geoscience, v. 3, no. 2, p. 251-267.

Godfrey, L.G., 1988, Misspecification tests in econometrics: The Lagrange Multiplier Principle and other approaches: New York, Cambridge University Press, 252 p.

Goldthwait, R.P., 1949, Artesian wells in New Hampshire: New Hampshire State Planning Development Commission, Mineral Resources Survey, part 11, 24 p.

Hatcher, R.D., Jr., Thomas, W.A., Veile, G.W., (eds.) 1989, The Appalachian-Ouachita Orogen in the United States: Boulder, Colo., Geological Society of America, The Geology of North America, v. F-2.

Hausman, J., 1978, Specification tests in econometrics: Econometrica, v. 46, p. 1,251-1,271.

Hsieh, P.A., Shapiro, A.M., Barton, C.C., Haeni, F.P., Johnson, C.D., Martin, C.W., Paillet, F.L., Winter, T.C., and Wright, D.L., 1993, Methods of characterizing fluid movement and chemical transport in fractured rock, in Chaney, J.T., and Hepburn, J.C., (eds.), 1993, Field trip guidebook for northeastern United States, Geological Society of America, Annual Meeting, Boston, Mass., October 25-28, 1993, p. R1-29.

Knopman, D.S., 1990, Factors related to the water-yielding potential of rocks in the Piedmont, Valley, and Ridge Provinces of Pennsylvania: U.S. Geological Survey Water-Resources Investigations Report 90-4174, 52 p.

Knopman, D.S., Hollyday, E.F., 1993, Variation in specific capacity in fractured rocks, Pennsylvania: Groundwater, v. 31, no. 1, p. 135-145.

Lattman, L.H., 1958, Technique of mapping geologic fracture traces and lineaments on aerial photographs: Photogrammetric Engineering, v. 24, p. 568-576.

LeGrand, H.E., 1967, Groundwater of the Piedmont and Blue Ridge Provinces in the southeastern states: U.S. Geological Survey Circular 538, 11 p.

Lyons, J.B., Bothner, W.A., Moench, R.H., and Thompson, J.B., Jr., 1997, Bedrock geologic map of New Hampshire: U.S. Geological Survey State Geologic Map, 2 sheets, scale 1: 250,000 and 1:500,000.

Mabee, S.B., 1992, Lineaments: Their value in assessing groundwater availability and quality in bedrock aquifers of glaciated metamorphic terrains—A case study: Amherst, Mass., University of Massachusetts, Ph.D. dissertation, 567 p.

———1996, Effectiveness of remotely-sensed lineaments and outcrop-scale fractures in identifying bedrock aquifers in New England: Amherst, Mass., University of Massachusetts Grant Proposal No. 96A1281, 28 p.

Mabee, S.B., and Hardcastle, K.C., 1997, Analyzing outcrop-scale fracture features to supplement investigations of bedrock aquifers: Hydrogeology Journal, v. 5, no. 4, p. 21-36.

Mabee, S.B., Hardcastle, K.C., and Wise, D.U., 1994, A method of collecting and analyzing lineaments for regional-scale fractured-bedrock aquifer studies: Groundwater, v. 32, no. 6, p. 884-894.

Medalie, Laura, and Moore, R.B., 1995, Ground-water resources in New Hampshire: Stratified-drift aquifers: U.S. Geological Survey Water-Resources Investigations Report 95-4100, 31 p.

Moore, R.B., and Clark, S.F., Jr., 1995, Assessment of ground-water supply potential of bedrock in New Hampshire: U.S. Geological Survey Ground-Water Supply Fact Sheet, FS 95-002, 2 p.

Moore, R.B., Clark, S.F., Jr., Ferguson, E.W., Marcoux, G.J., and Degnan, J.R., 1998, New Hampshire bedrock aquifer assessment: Correlating yields from 18,000 wells with geologic features, *in* Northeast Focus Ground Water Conference, Burlington, Vt., October 20-21, 1998, Proceedings: Burlington, Vt., National Ground Water Association, p. 18.

Moore, R.B., Flanagan, S.F., Cotton, J.E., Ayotte, J.D., Olimpio, J.R., Chormann, F.H., Jr., 1999, Stratified-drift aquifers in New Hampshire, U.S. Geological Survey outreach-poster map, 1:250,000 scale.

Moore, R.B., Schwarz, G.E., Clark, S.F., Jr., Walsh, G.J., Degnan, J.R., and Mack, T.J., 2001, Relating well yields to site characteristics in fractured bedrock of New Hampshire: Abstracts with Programs—Geological Society of America, *in* Geological Society of America, 2001 Annual Meeting, November 2001.

National Research Council, 1996, Rock fractures and fluid flow, contemporary understanding and applications: Washington, D.C., National Academy Press, 551 p.

Pohn, H.A., 1987, Radar and landsat lineament maps of the Glens Falls 1° by 2° quadrangle; New York, Vermont, and New Hampshire: U.S. Geological Survey Open-File Report 86-471, 2 sheets, scale 1:250,000, 7 p.

Salvini, Francesco, Billi, Andrea, and Wise, D.U., 1999, Strike-slip fault propagation cleavage in carbonate rocks: the Mattinata Fault Zone, Southern Apennines, Italy: Journal of Structural Geology, v. 21, p. 1,731-1,749.

Sander, Per, 1997, Water-well siting in hard-rock areas: Identifying promising targets using a probabilistic approach: Hydrogeology Journal, v. 5, no. 3, p. 32-43.

Sander, Per, Minor, T.B., and Chesley, M.M., 1997, Ground-water exploration based on lineament analysis and reproducibility tests: Groundwater, v. 35, p. 888-894.

Seeburger, D.A., and Zoback, M.D., 1982, The distribution of natural fractures and joints at depth in crystalline rock: Journal of Geophysical Research, v. 87, no. B7, p. 5,517-5,534.

Shapiro, A.M., and Hsieh, P.A., 1991, Research in fractured rock hydrogeology: Characterizing fluid movement and chemical transport in fractured rock at the Mirror Lake drainage basin, New Hampshire, *in* Mallard, G.E., and Aronson, D.A., (eds.), Toxic Substances Hydrology Program, Technical Meeting, Monterey, Calif., March 11-15, 1991, Proceedings: U.S. Geological Survey Water-Resources Investigations Report 91-4034, p. 155-161.

Siddiqui, S.H., 1969, Hydrogeologic factors influencing well yields and hydraulic properties of folded and faulted carbonate rocks in central Pennsylvania: Pennsylvania State University, unpublished Ph.D. dissertation, 502 p.

Siddiqui, S.H., and Parizek, R.R., 1971, Hydrogeologic factors influencing well yields in folded and faulted carbonate rocks in central Pennsylvania: Water Resources Research, v. 7, no. 5, p. 1,295-1,312.

———1972, Application of nonparametric statistical tests in hydrology: Groundwater, v. 10, p. 26-31.

———1974, An application of parametric statistical tests to well-yield data from carbonates of central Pennsylvania: Journal of Hydrology, v. 21, p. 1-14.

Sriramadas, Aluru, 1962, The geology of the Manchester quadrangle, New Hampshire: Concord, N.H., New Hampshire Department of Resources and Economic Development Bulletin, v. 2, 78 p.

Stewart, G.W., 1968, Drilled water wells in New Hampshire: Concord, N.H., New Hampshire Department of Resources and Economic Development, 58 p.

Swanson, M.T., 1988, Pseudotachylyte-bearing strike-slip duplex structures in the Fort Foster brittle zone of southernmost Maine: Journal of Structural Geology, v. 10, p. 813-828.

U.S. Geological Survey, 1984, National Water Summary 1984—Hydrologic events, selected water-quality trends, and ground-water resources: Water-Supply Paper 2275, 467 p.

Walsh, G.J., and Clark, S.F., Jr., 1999, Bedrock geologic map of the Windham quadrangle, Rockingham and Hillsborough Counties, New Hampshire: U.S. Geological Survey Open-File Report 99-8, scale 1:24,000, 19 p.

———2000, Contrasting methods of fracture trend characterization in crystalline metamorphic and igneous rocks of the Windham quadrangle, New Hampshire: Northeastern Geology and Environmental Sciences, v. 22, no. 2, p. 109-120.

White, H., 1982, Maximum likelihood estimation of misspecified models: Econometrica, v. 50, p. 1-25.

Wise, D.U., Funiciello, R., Parotto, M., and Salvini, F., 1985, Topographic lineament swarms: Clues to their origin from domain analysis of Italy: Geological Society of America Bulletin, v. 96, p. 952-967.

APPENDIX A

APPENDIX A. REGRESSION MODEL CONSTRUCTION

The overall approach in deciding which variables to include in the model was as follows:

1. Ordinary least-squares regression was used to provide an initial estimate of which variables are significant predictors of well depth for the first stage of the model. Significant predictors of the natural log of depth of the well then could be identified.
2. These predictor variables then were used as the starting point for building the two-stage model. This process favors the inclusion of variables that affect well depth and yield.
3. Predictor variables were removed from the two-stage model if they were insignificant as predictors of the natural log of yield.
4. The two-stage model was built by comparing the residuals with the remaining unused variables. Guided by the relation, new variables were introduced into the two-stage model. The major groups of lithologies, which cover all of New Hampshire, had priority over individual map units to make a more inclusive model.
5. Regional variables were tested in the model, which created a nested model. Statewide mapped bedrock units were regionally subdivided into more detailed categories. The more general statewide variables remained active in those regions (Pinardville and Windham quadrangles) where detailed mapping was conducted.
6. Additional variables tested in the model included uphill drainage area, proximity to surface water, driller (as an instrumental variable), construction method, and median household income (also as an instrumental variable).

APPENDIX B

Appendix B. List of variables tested for the regression model

[DEM, Digital elevation model; DLG, Digital line graph]

Variables	Definition of variable	* Indicator variable
	DEPENDENT VARIABLES	
YIELD	Yield, in gallons per minute	
lnYIELD	Natural log of yield	
YIELDFT	Yield/length, in gallons per minute per foot of open hole	
	PREDICTOR VARIABLES	
lnDEPTH	Natural log of the well depth, in feet	
OVER	Overburden depth to bedrock	
OPEN	Length of open hole	
CSN	Casing length	
IPLUTN	Dichotomous variable identifying wells drilled into plutons	*
Seven major lithologic units		
LITHO1	Lithologic unit 1: medium to coarse-grained (nonfoliated to weakly foliated igneous rock)	*
LITHO2	Lithologic unit 2: well-foliated, coarse-grained igneous rock	*
LITHO3	Lithologic unit 3: fine-grained, tightly bonded metasedimentary rock	*
LITHO4	Lithologic unit 4: fine to medium-grained granular, metasedimentary rock;	*
LITHO5	Lithologic unit 5: graphitic micaceous metapelite, locally interbedded with calcareous or siliceous rock	*
LITHO6	Lithologic unit 6: fine and coarse-grained, volcanic rock and associated metapelite	*
LITHO7	Lithologic unit 7: Massabesic Gneiss Complex and Breakfast Hill Granite	*
All bedrock units, with well data, are shown on the 1997 State Bedrock Geologic Map (Lyons and others, 1977)		
	Bedrock unit	
RKTYP2	Ꞓjb	*
RKTYP3	Dir	*
RKTYP4	Dif	*
RKTYP5	K1a	*
RKTYP6	K1bx	*
RKTYP7	K1r	*
RKTYP8	K2	*
RKTYP9	K7C	*
RKTYP14	Kc1b	*
RKTYP21	J1r	*
RKTYP22	J1x	*
RKTYP23	J7h	*
RKTYP25	J4hx	*
RKTYP27	J5	*
RKTYP28	Sfrv	*
RKTYP29	Oalx	*
RKTYP30	J7x	*
RKTYP31	J8	*

Appendix B. List of variables tested for the regression model—Continued

Variables	Definition of variable	* Indicator variable
	Bedrock unit--Continued	
RKTYP32	J9A	*
RKTYP35	Jc1b	*
RKTYP37	Jo1b	*
RKTYP38	Jo1h	*
RKTYP39	P1m	*
RKTYP40	Oalb	*
RKTYP43	D1b	*
RKTYP44	Oalg	*
RKTYP45	D1m	*
RKTYP49	Oh1-2h	*
RKTYP50	D3Bb	*
RKTYP52	Sfrx	*
RKTYP53	DS9	*
RKTYP54	DSlr	*
RKTYP55	Db2b	*
RKTYP57	Dc1m	*
RKTYP58	Dg	*
RKTYP59	Dgc	*
RKTYP60	Oh2-9A	*
RKTYP61	Dgm	*
RKTYP62	Dgv	*
RKTYP63	Di	*
RKTYP64	Dih	*
RKTYP65	Dk2x	*
RKTYP66	Dl	*
RKTYP67	Dlc	*
RKTYP69	Oo4Ch	*
RKTYP70	Dll	*
RKTYP71	Dlu	*
RKTYP72	Dlv	*
RKTYP74	Oalf	*
RKTYP75	Oals	*
RKTYP77	Ds1-6	*
RKTYP78	Ds6-9B	*
RKTYP79	D3Ab	*
RKTYP80	Oaux	*
RKTYP81	PM1m	*
RKTYP82	K4x	*
RKTYP83	Dw3A	*

Appendix B. List of variables tested for the regression model—Continued

Variables	Definition of variable	* Indicator variable
	Bedrock unit--Continued	
RKTYP86	Sa2x	*
RKTYP87	Sc	*
RKTYP88	Sf	*
RKTYP89	Sfc	*
RKTYP90	Sfr	*
RKTYP91	Sfrb	*
RKTYP92	Sfrc	*
RKTYP93	Sfrg	*
RKTYP95	Sg	*
RKTYP96	SObg	*
RKTYP98	Smsf	*
RKTYP99	SOb	*
RKTYP100	Dc3Am	*
RKTYP101	Sp	*
RKTYP102	Spr	*
RKTYP103	SOe	*
RKTYP104	Spvs	*
RKTYP105	Spvx	*
RKTYP106	Sr	*
RKTYP107	SOk	*
RKTYP108	Src	*
RKTYP109	Srl	*
RKTYP110	Srlp	*
RKTYP111	Sru	*
RKTYP112	Srup	*
RKTYP114	Ssf	*
RKTYP115	Ssfb	*
RKTYP118	De9	*
RKTYP119	Ssfx	*
RKTYP120	Zmz	*
RKTYP121	SObc	*
RKTYP123	Sn2-3A	*
RKTYP124	Sn1x	*
RKTYP128	Oaus	*
RKTYP136	Oh2h	*
RKTYP139	Oo1-3A	*
RKTYP140	Oo1-3B	*
RKTYP142	Oo1b	*
RKTYP144	Oo1h	*

Appendix B. List of variables tested for the regression model—Continued

[DEM Digital ele ation model; DLG, Digital line gra h]

Variables	Definition of variable	* Indicator variable
	Bedrock unit--Continued	
RKTYP145	Oo2-3A	*
RKTYP146	Oo2b	*
RKTYP147	Oo2bx	*
RKTYP148	Oo2h	*
RKTYP149	Oo3B	*
RKTYP150	Oo3B-6	*
RKTYP152	Oo4-7h	*
RKTYP159	Opv	*
RKTYP160	Oq	*
RKTYP161	O-Cd	*
RKTYP162	O-Cdp	*
RKTYP163	O-Czl	*
RKTYP164	O-Czu	*
RKTYP173	OZrb	*
RKTYP174	OZrz	*
Topographic variables		
SLOPE	Percent slope derived from 1:24,000-scale DEM	
SECDRV	Curvature derived from 1:24,000-scale DEM	
SLOPE250	Slope derived from 1:250,000-scale DEM	
CURV250	Curvature derived from 1:250,000-scale DEM	
HILLTOP	Derived from 1:24,000-scale DEM	*
HILLTOP2	Derived from 1:250,000-scale DEM	*
DRAW2	Valley bottom or swale derived from 1:250,000-scale DEM	*
DRAW3	Valley bottom or swale derived from 1:24,000-scale DEM	*
HILLSIDE	Derived from 1:24,000-scale DEM	*
HILLSID2	Derived from 1:250,000-scale DEM	*
ELEV24	Elevation derived from 24:000-scale map or DEM	
lnACCUM	Natural log of the number of DEM grid cell uphill of the well	
ACCUM	The number of DEM grid cell uphill of the well	
DISTH20	Distance to surface water, in feet determined from the 1:24,000-scale DLG	
LNDSTH20	Natural log of the distance to surface water	
SDCODE	Stratified drift	*
Proximity to lineaments (within 100 feet)		
DSTLO	Proximity to low-altitude lineaments	*
DSTLSAT	Proximity to Landsat lineaments	*
DSTHI	Proximity to high-altitude lineaments	*
DSTALL3	Proximity to all three	*

Appendix B. List of variables tested for the regression model—Continued

Variables	Definition of variable	* Indicator variable
Interactions between major lithologic groupings and proximity to lineaments		
L1HI L2HI L3HI L4HI L5HI L6HI		*
L7HI L1LO L2LO L3LO L4LO L5LO L6LO L7LO		*
L1LSAT L2LSAT L3LSAT L4LSAT L5LSAT		*
L6LSAT L7LSAT R29HI		*
Interactions between plutons and lineaments		
PLUTHI		*
PLUTLO		*
PLUTLSAT		*
PLUTALL3		*
Interactions between metasediments and lineaments		
METAHI		*
METALO		*
METALSAT		*
METALL3		*
Interaction between plutons, lineaments, and 10-degree orientation categories		
PAC1-PAC18		*
Interaction between metasediments, lineaments, and 10-degree orientation categories		
MAC1-MAC18		*
Interactions between rock units and proximity to lineaments		
R29LO R29LSAT R35HI R35LO R35LSAT R38HI R38LO R38LSAT R39HI R39LO		*
R39LSAT R45HI R45LO R45LSAT R55HI R55LSAT R57HI R57LO R57LSAT		*
R66HI R66LO R66LSAT R71HI R71LO R71LSAT R77HI R77LO R77LSAT R81HI		*
R81LO R83HI R83LO R83LSAT R86HI R86LO R86LSAT R97HI R97LO		*
R97LSAT R101HI R101LO R101LSAT R102HI R102LO R102LSAT R103HI R103LO		*
R103LSAT R107HI R107LO R107LSAT R114HI R114LO R114LSAT R118HI R118LO		*
R118LSAT R120HI R120LO R120LSAT R121HI R121LO R121LSAT R122HI		*
R122LO R122LSAT ...		*
Interactions between major lithologic groupings and topographic variables		
L1SDRV L2SDRV L3SDRV L4SDRV L5SDRV L6SDRV L7SDRV		*
L1HILL L2HILL L3HILL		*
L4HILL L5HILL L6HILL L7HILL L1HLL2 L2HLL2 L3HLL2 L4HLL2 L5HLL2 L6HLL2 L7HLL2		*
L1DRAW L2DRAW L3DRAW L4DRAW L5DRAW L6DRAW L7DRAW		*
L1DRAW3 L2DRAW3 L3DRAW3 L4DRAW3 L5DRAW3 L6DRAW3 L7DRAW3		*
Water use		
COMERCL	Commercial supply well	*
PUBLIC	Public supply well	*
Construction method		
CABLE	Well drilled with a cable tool rig	*

Appendix B. List of variables tested for the regression model—Continued

Variables	Definition of variable	* Indicator variable
Demand instruments		
YEAR	Year well drilled	*
INCOME	Median household income from 1990 census data	*
DRILLER1-DRILLER25	Drillers in the State with the most number of wells drilled	*
	REGIONAL VARIABLES	
Lineaments (The following lineament variables restricted to just plutons or just metasediments)		
TOPOBUF	Within 100 feet of a lineament identified by the use of the topographic contours	*
CIRBUF	Within 100 feet of a lineament identified by the use of color infrared photographs	*
Fracture-correlated lineaments - DOMAIN ANALYSIS (primary fracture set)		
D180BUF	Within 100 feet of a 80,000 lineament -- domain analysis	*
D120BUF	Within 100 feet of a 20,000 lineament -- domain analysis	*
D1LSTBUF	Within 100 feet of a Landsat lineament -- domain analysis	*
D1TPOBUF	Within 100 feet of a topographic lineament -- domain analysis	*
D1CIRBUF	Within 100 feet of a color infrared lineament -- domain analysis	*
Fracture-correlated lineaments - DOMAIN ANALYSIS (any peak greater than 30 percent of primary peak)		
D380BUF	Within 100 feet of a 80,000 lineament -- domain analysis	*
D320BUF	Within 100 feet of a 20,000 lineament -- domain analysis	*
D3LSTBUF	Within 100 feet of a Landsat lineament -- domain analysis	*
D3TPOBUF	Within 100 feet of a topographic lineament -- domain analysis	*
D3CIRBUF	Within 100 feet of a color infrared lineament -- domain analysis	*
Fracture-correlated lineaments - DOMAIN ANALYSIS (any peak greater than 50 percent of primary peak)		
D580BUF	Within 100 feet of a 80,000 lineament -- domain analysis	*
D520BUF	Within 100 feet of a 20,000 lineament -- domain analysis	*
D5LSTBUF	Within 100 feet of a Landsat lineament -- domain analysis	*
D5TPOBUF	Within 100 feet of a topographic lineament -- domain analysis	*
D5CIRBUF	Within 100 feet of a color infrared lineament -- domain analysis	*
Fracture-correlated lineaments - 1,000 foot buffer technique		
K80FRACT	High-altitude lineament	*
K20FRACT	Low-altitude lineament	*
LSTFRACT	Landsat lineament	*
TPOFRACT	Topographic map lineament	*
CIRFRACT	Color infrared lineament	*
ALLFRACT	High-altitude, Landsat, topographic, or color infrared lineament	*

Appendix B. List of variables tested for the regression model—Continued

Variables	Bedrock unit	Definition of variable	* Indicator variable
Geologic units mapped in the Pinardville and Windham quadrangles, New Hampshire			
DG	Dg	Binary granite and granitic pegmatite	*
DMG	Dmg	Muscovite granite	*
DP	Dp	Pegmatite	*
DSG	Dsg	Spaulding granite	*
MD	Md	Diabase and lamprophyre dike	
MDP	Mdp	Porphyritic diabase dike	*
PDMG	PDmg	Migmatite gneiss	*
PZMG	PZmg	Layered migmatite and restite	*
PDG	Pdg	Damon Pond granite	*
PPG	Ppg	Granite and pegmatite	*
SOB	SOb	Biotite-plagioclase-quartz granofels and schist and calc-silicate rock	*
SOBQ	SObq	Quartzite, biotite-plagioclase-quartz granofels, and calc-silicate rock	*
SOBW	SObw	Well bedded biotite-plagioclase-quartz granofels and schist and calc-silicate rock	*
SAG	Sag	Porphyritic granite to granodiorite	*
SAGD	Sagd	Granodiorite	*
SAGDG	Sagdg	Garnetiferous granodiorite	*
SR	Sr	Light-gray biotite-muscovite schist	*
SRB	Srb	Biotite schist and granofels	*
SRLB	Srlb	Layered biotite granofels	*
SRMG	Srmg	Migmatite schist and gneiss	*
SRQ	Srq	Rusty quartzite and schist	*
SRR	Srr	Rusty-biotite-muscovite schist	*
ZMA	Zma	Amphibolite gneiss	*
ZMLG	Zmlg	Layered paragneiss and orthogneiss	*
CNTXBUF		Within 100 feet of a mapped contact	*
AXMBUF		Within 100 feet of a mapped fold axis	*
DIKBUF		Within 100 feet of a mapped dike	*
FAUBUF		Within 100 feet of a mapped fault	*
ISOBUF		Within 100 feet of a mapped isograd	*
AXMBUF		Within 2,000 feet of the axis of the Massabesic Gneiss Complex anticline	*

REFERENCE CITED

Lyons, J.B., Bothner, W.A., Moench, R.H., and Thompson, J.B., Jr., 1997, Bedrock geologic map of New Hampshire: U.S. Geological Survey State Geologic Map, 2 sheets, scale 1: 250,000 and 1:500,000.

APPENDIX C

APPENDIX C. REGRESSION MODEL TESTS

The statewide analysis of bedrock yields by this project involved developing a complex statistical model. Large data sets were used to examine geospacial data, and the resultant model was subjected to a variety of statistical tests.

Test for an Endogenous Variable

A test was applied to evaluate the use of the instrumental variable technique to eliminate the effect of the inherent relation between well yield and depth. The use of instrumental variables is not a preferred method of estimation if the predictor variables are not correlated with the error term. Thus, in order to determine the appropriateness of this method, a statistical test of the predictor variable, natural logarithm of depth, was required to determine if it is an endogenous variable. The test used is described by Hausman (1978) and evaluates the null hypothesis that the regressors and the residual are independent. Under the null hypothesis, coefficient estimates obtained by either ordinary least squares (OLS), or by instrumental variables, are consistent and unbiased in large samples. Also, under the null hypothesis, the method of OLS produces more precise coefficient estimates than do instrumental variables. The test consists of obtaining two sets of estimates of the coefficients, one estimated by OLS and the other estimated using instrumental variables, and determining if the difference in the coefficient estimates is significantly different from zero.

An "omitted variables" form of the test is described by Godfrey (1988, p.194-195). The original regression equation is augmented with additional explanatory variables, which consist of the residuals of the potentially endogenous variable (lnDEPTH) determined from an OLS regression of lnDEPTH on the original set of instrumental variables not under investigation. The augmented equation is estimated using the instrumental-variable technique with instruments consisting of the original set, plus the potentially endogenous variables. If the coefficients associated with the augmented variables are jointly significant, as determined by a likelihood-ratio-test statistic distributed as chisquare with degrees of freedom equal to the number of augmented variables (1), then the null hypothesis that the potentially endogenous variables are exogenous and can be rejected.

Applying the test described in the previous paragraph, lnDEPTH is shown to be endogenous, and the use of a two-stage model with instrumental variables is appropriate. Results of the test yield a chi-square test statistic of 521 with one degree of freedom, and a significance of 0.00. As the significance is less than 0.05, the null hypothesis that lnDEPTH is exogenous was rejected.

Specification Test

A "well-specified" model (Godfrey, 1988) is one where the functional form is correct and the residuals are homoscedastic (for example, the distribution of the residuals are similar throughout the range of predicted values). The coefficient estimates and the associated estimated standard errors are unbiased. Concurrently, predictions from the model will be unbiased and the standard formula for evaluating prediction accuracy will be correct.

There are many tests that can be used to validate model specification. Godfrey (1988, chap. 4) presents a comprehensive description of specification tests. The test used (White, 1982; p. 1-25) for the regression model addresses the issue of specification if there is no explicit alternative model under consideration. The test determines if squared values of residuals are correlated with instrumental variables. The test is sensitive to misspecification in the regression equation and heteroscedasticity in the data and will likely trigger a false positive indication of misspecification, particularly if there are a lot of instrumental variables used.

The test is significant for the model, as determined by its F statistic, indicating that there is evidence of heteroscedasticity. The F statistic is 194.5 with a p-value less than 0.0001. The null hypothesis that there is no misspecification is rejected. As most predictors are dichotomous, the likely cause of the rejection is heteroscedasticity. All nine predictors of lnYIELD, which were identified as significant by the test (table C1), are dichotomous. In other words, there are significantly different variances between populations where the indicator variables are 1 and where they are 0. The weight column in table C1 represents the ratio of the variance of the residuals, where the variable is zero, divided by the variance of the residuals, where the variable is one.

Table C1. Significant predictors in the specification test of the model used in this study

[All variables are defined in table 2]

Variable	Coefficient	Significance	Weight ratio
SOb	0.45	0.0001	0.82
Zmlg	-.45	.0199	.95
PZmg	-1.13	.0180	1.49
RKTYP103	-.40	.0291	1.22
RKTYP107	-.64	.0101	1.39
RKTYP120	.69	.0001	.79
RKTYP57	-.25	.0165	1.16
CABLE	-1.06	.0001	2.33
LITHO2	.15	.0203	1.00

The practical implications of the tests described previously are that estimated probabilities will be somewhat over or underestimated for sites in the following areas. For those sites where the weight ratio is less than one (SOb, Zmlg, and RKTYP120), the probability of obtaining a given yield tends to be underestimated for probabilities less than 50 percent and overestimated for probabilities greater than 50 percent. For those sites where the weight is greater than one (PZmg, RKTYP103, RKTYP107, and RKTYP57), probabilities tend to be overestimated for probabilities less than 50 percent and underestimated for probabilities greater than 50 percent. The weight for LITHO2 is close to 1 and the probability estimates should be consistent if LITHO2 is 0 or 1. There also is much less variance among wells drilled by cable tool than for wells drilled using other techniques. In predicting probabilities, however, it is assumed that the site is not to be drilled by cable-tool construction, and an adjustment of the probabilities is not necessary.

Residual Statistical Distribution

Testing for normality of the residuals is important with a regression model. The cumulative distribution function of the residuals was determined by use of SAS/INSIGHT software. Confidence bands, at the 95-percent confidence level, also were determined using SAS/INSIGHT. The Kolmogorov statistic D, the maximum vertical distance between the two distribution functions, was then used to test the null hypothesis that the population distribution is normally distributed. In the normality test, the null hypothesis is rejected because at some point, the normal distribution falls outside the confidence band (fig. C1).

Although the Kolmogorov statistic D test indicates that the residuals are not normally distributed, at the 95-percent confidence level, the graphical distribution (fig. C1 (A and B)) is near normal. Thus, for practical purposes, the model is considered appropriate for estimating probabilities equaling or exceeding a given yield.

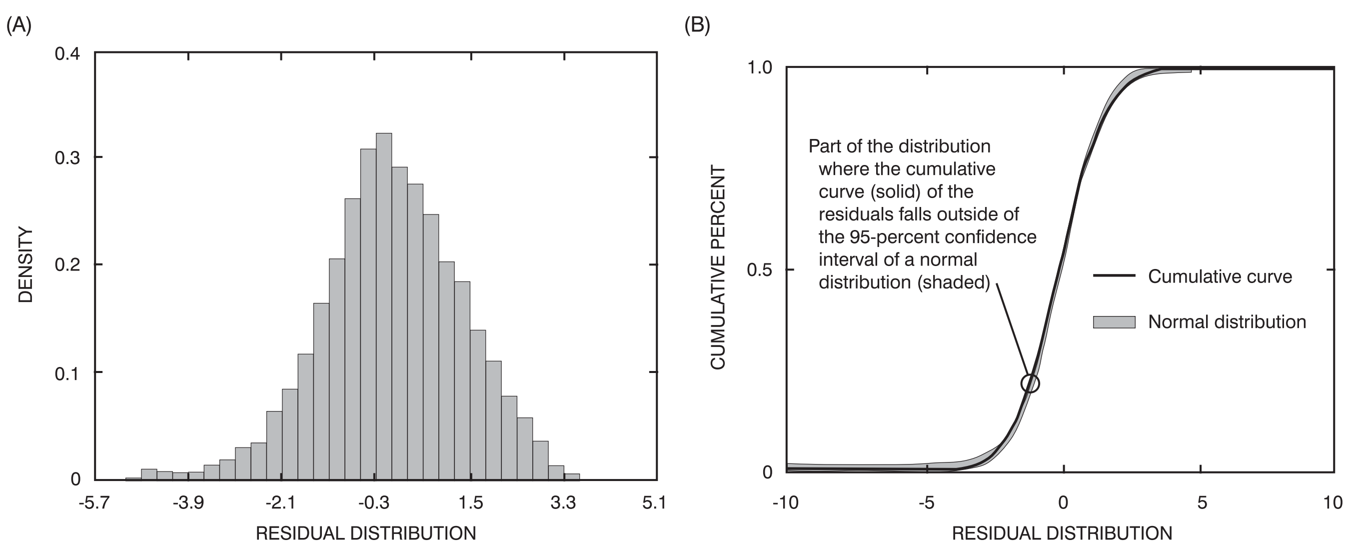

Figure C1. Distribution of model residuals for New Hampshire shown as a (A) bell curve and (B) cumulative curve. Kolmogorov statistic D is 0.0209 with a probability greater than D of less than 0.01.

Residual Spatial Distribution and Spatial Analysis

The spatial distribution of residuals was evaluated to see if there is a need for further regionalization of the model. Statewide maps of the residuals were made to see if positive or negative residuals (color coded) are grouped in regions. A non-uniform coverage of well distribution can cause bias in a misspecified regression model, and there are areas in southern New Hampshire with significantly higher densities of data than in the rest of the State. The regression model, however, accommodates this by using predictor variables that describe the well-site characteristics. Areas with large amounts of data in southern New Hampshire potentially could dominate the equation. If this happened, then a pattern should be apparent on statewide plots of the residuals maps. Areas with few data would show a bias positively or negatively, and this was not observed in the map. Instead, an even distribution of positive and negative residuals generally were displayed.

Locally, spatial distributions also were evaluated in the Pinardville quadrangle where numerous wells are clustered. There is a similarity of well yields with medium and low yields (less than 40 gal/min) for wells aligned in a direction north-northeast to one another among and in specific geologic units (Drew and others, 1999). The hypothesis of the model described in this report, however, is that spatial features inherent in the predictor variables explain systematic spatial features of the yield data. If this hypothesis is correct, then the residuals from the regression model will be independent. If it is not correct, then standard errors of the regression coefficients will be biased downward and predictions of well yield at a given site could be improved by incorporating known yield data at neighboring sites. This result can be tested by examining the spatial distribution of the model residuals.

Declustering of the data was not necessary, on the basis of a spatial test of the residuals for the Pinardville 7.5-minute quadrangle (fig. C2). Pinardville is the quadrangle with the highest density of data in the State. Results of the test did not show spatial significance, therefore, corrections to the model for this potential bias are not considered necessary.

Estimates of spatial correlation for model residuals in the Pinardville 7.5-minute quadrangle are shown in figure C2. Correlations are computed by binning the separation distances between pairs of wells. The results for binning separation distances into 100 cells across the quadrangle are shown in figure C2. The results of the estimates show little evidence of spatial correlation or of any directional preference in residual correlations. All correlations are small even at fine scales such as a 1,000-cell analysis (not shown).

Numerical correlations for the center of the separation space are listed in table C2. For the 100-cell analysis, correlations are all statistically insignificant. Each correlation in the table is based on at least 1,332 pairs. There are few correlations that exceed the significance threshold (or absolute value of the correlation greater than 1 over the square root of n, where n is the number of paired residuals used to compute the correlation for a given cell—the minimum n is 337). There is no recognizable pattern to the correlogram (fig. C2), and the statistical significance of certain correlations appears random. A 1,000-cell analysis (not shown) has fewer observations per cell than a 100-cell range. Again, none of these results showed evidence of a spatial pattern. The site characteristics that are used as input to the regression model, including the additional regional variables, account for the spatial relation observed between well yields; therefore, the spatial correlation of the model residuals is not a concern.

Comparison of Verification and Primary Data Sets

There are a number of tests available for comparing the results of a primary data set with those of a verification data set. Chow (1960) provides a test to determine if there is a significant difference in the coefficient estimates of two independent regressions. Two sets of data combined into a common data set. A test model is then constructed in which each coefficient in the model is estimated with a companion coefficient that is multiplied by a verification data-set indicator. For example, the effect of depth in the test model would be represented as

$$(b + dI)\ln(DEPTH), \qquad (3)$$

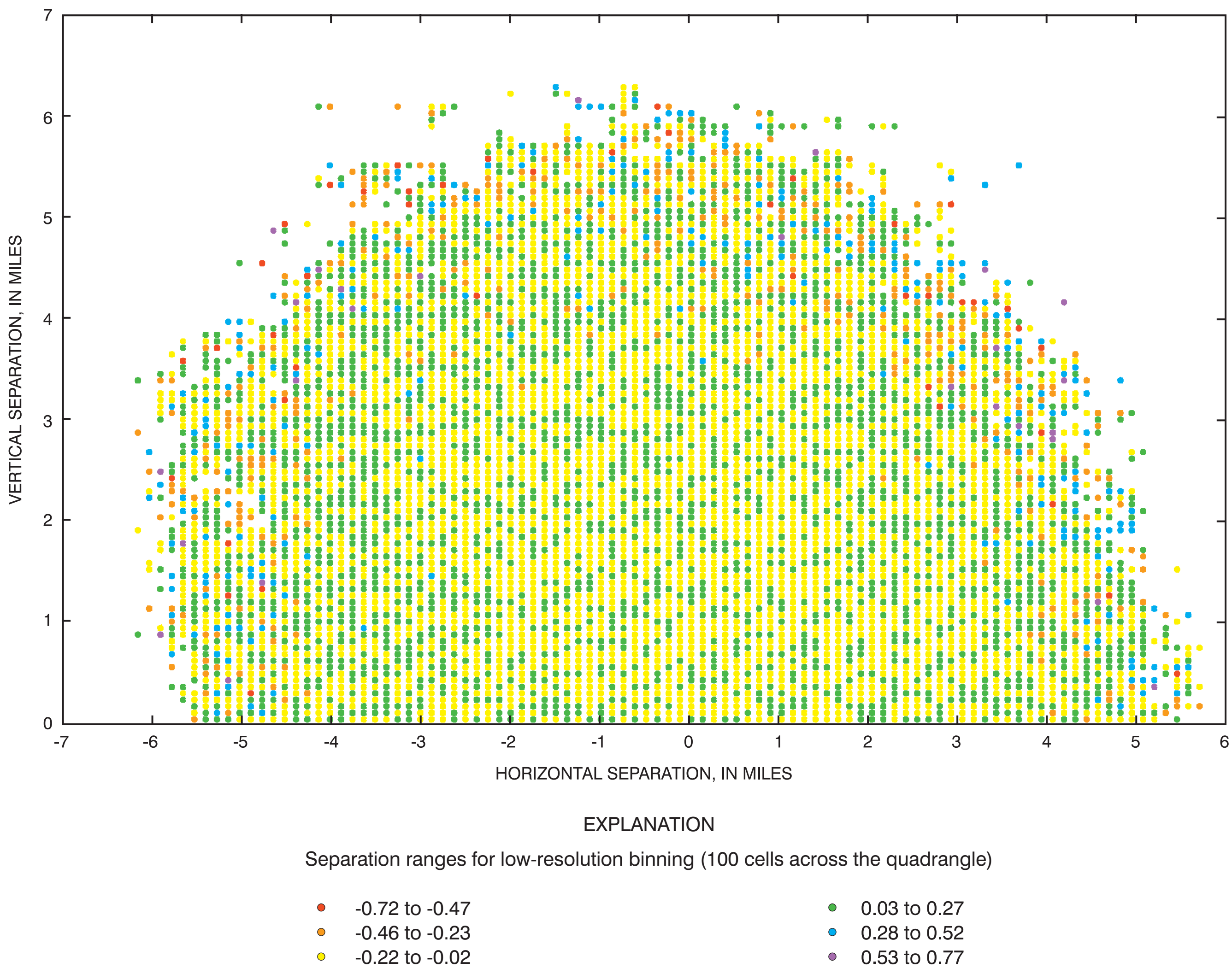

Figure C2. A correlogram showing the spatial correlation of model residuals that have a low-resolution binning (100 cells across) for the Pinardville quadrangle, New Hampshire. A spatial correlation pattern is not apparent.

Table C2. Numerical correlations of the model residuals for the Pinardville quadrangle, New Hampshire, for distance separations less than 1 kilometer

[The correlations in greater detail for x and y spatial separations less than 1 kilometer (km); * = correlations that exceed the significance threshold (absolute value of the correlation greater than 1 over the square root of n, where n is the number of paired residuals used to compute the correlation for a given cell; the minimum n is 337)]

Vertical distance separation (km)	**Horizontal separation (km)**									
	-0.91	-0.71	-0.50	-0.30	-0.10	0.10	0.30	0.50	0.71	0.91
0.98	0.00	-0.07	0.01	-0.05	0.08	0.01	0.02	0.07	0.01	-0.03
.88	.07	-.02	.00	.06	-.05	.06	-.03	.04	-.01	-.02
.77	-.04	-.04	.08	-.01	-.02	.03	.01	.00	.03	-.07
.67	-.05	-.04	.06	.00	-.04	.03	.05	.05	.08	-.04
.57	.03	.04	.02	-.00	.05	-.01	.04	-.02	-.05	.00
.46	.02	-.06	-.06	-.06	-.03	-.07	.01	.01	-.03	.09
.36	-.00	.01	.06	-.11 *	-.01	-.01	.05	.03	-.02	-.00
.26	.04	.12 *	.02	.07	-.05	.00	.02	.01	-.05	-.01
.15	-.06	.01	-.03	-.03	-.01	.02	-.00	-.00	-.03	-.04
.05	.03	-.02	.03	.01	-.00	.07 *	.04	-.03	.09	.16 *

where

- b is the depth coefficient,
- d is the depth companion coefficient,
- I is the indicator variable for the verification data set (I equals 1 if the observation is from the verification data set and equals zero otherwise), and
- *lnDEPTH* is the regressor natural log (ln) of depth.

The test rejects the null hypothesis that the results are consistent across data sets if the companion coefficients are jointly significant. This rejection is determined by a likelihood-ratio statistic, in which distributed chi-square with degrees of freedom are equal to the number of companion coefficients estimated. Coefficients for only two of the predictor variables were found to be jointly significant at the 95-percent confidence level when applying the above test to the pooled data set. These two variables are PZmg (layered migmatite of the Massabesic Gneiss Complex), and RKTYP72 (Dlv, Littleton formation with volcanic lentils). With 44 variables, and a 95-percent confidence level, two or three variables should test significant randomly.

Using a summary Chow test, however, to compare the two subsamples, indicates that there is a statistically significant difference between the verification and the primary data sets. The computed Chow test statistic is 64.46, with 45 degrees of freedom, and a significance level of 0.030. The null hypothesis that the subsamples are the same is rejected because the significance level is less than 0.05. The variable, which accounts for the largest difference between the sequestered data and the primary data set is PZmg (with coefficients of opposite sign between the two data sets). Eliminating this variable from the model would decrease the Chow statistic to 39.90 with a significance of 0.65, in which case the null hypothesis that the subsamples are the same would be accepted.

Finite Bias Test

The model also is tested for finite-sample bias. Problems may be present with instrumental-variables estimation if the relation between the instruments and the endogenous variable are weak (Bound and others, 1995). According to Bound, weak instruments may lead to bias in the coefficients even if there is a large sample such as we have in this study. The bias is in the same direction as the bias under ordinary least squares (for example, the coefficient for lnDEPTH would be less than its true value). To test for this bias, an F statistic for the significance of the "excluded" instruments, the year, income, and driller variables is computed. This statistic is equal to

$$F = ((n - k)/q)(R^2 - R_r^2)/(1 - R^2), \qquad (4)$$

where

- n is the number of observations (16,302),
- k is the number of instruments for the unrestricted regression (for example, the total number of instruments) (70)
- q is the number of "excluded" instruments (for example, the number of instruments not used in the second stage of the model) (23),
- R^2 is the r-square for the unrestricted regression (the R-square for the first-stage regression of depth on all the instruments including the public supply and commercial indicator variables and all the other exogenous variables that appear in the yield equation) (0.202259), and
- R^2_r is the r-square for the restricted regression (the R-square for a regression of depth on all the exogenous variables in the yield equation but not the excluded variables) (0.148000).

Given these values, the value for the F statistic is 48.00. According to Bound and others (1995), 48.00 implies a finite sample bias of an instrumental variables coefficient estimate relative to the OLS bias of 0.00 (Bound and others, 1995, table A.1, $k = 2$ and t^2/k (= F) = 100). Thus, the test shows no indication of significant finite-sample bias.

Overall Statement

The model performs well given these test. As expected, the variable lnDEPTH was endogenous. Predicted probabilities may be over or underestimated in specific locations associated with certain heteroscedastic variables. However, the form of the model is appropriate. The residuals are near normally distributed, and spatial declustering of the data is not necessary. Also, the verification data set (with the possible exception of variable PZmg) indicates that the data set used to calibrate the model is composed of representative data. Lastly, there is no indication of significant finite-sample bias.

REFERENCES CITED

Bound, John, Jaeger, D.A., and Baker, R.M., 1995, Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak: Journal of the American Statistical Association, v. 90, no. 430, p. 443-50.

Chow, G.C., 1960, Tests of equality between sets of coefficients in two linear regressions: Econometrica, v. 28, p. 591-605.

Drew, L.J., Karlinger, M.R., Armstrong, T.R., and Moore, R.B., 1999, Relations between igneous and metamorphic rock, fracture patterns, and ground-water yield from the variography of water-well yields, Pinardville quadrangle, New Hampshire: Natural Resources Research, v. 8, no. 2, p. 137-152.

Godfrey, L.G., 1988, Misspecification tests in econometrics: The Lagrange Multiplier Principle and other approaches: New York, Cambridge University Press, 252 p.

Hausman, J., 1978, Specification tests in econometrics: Econometrica, v. 46, p. 1,251-1,271.

White, H., 1982, Maximum likelihood estimation of misspecified models: Econometrica, 50, p. 1-25.

This product was prepared by Ann Marie Squillacci (Layout and Design), Debra H. Foster (Editing), and Anita Cotton (Illustrations) in the U.S. Geological Survey New Hampshire/Vermont District, Pembroke, New Hampshire.