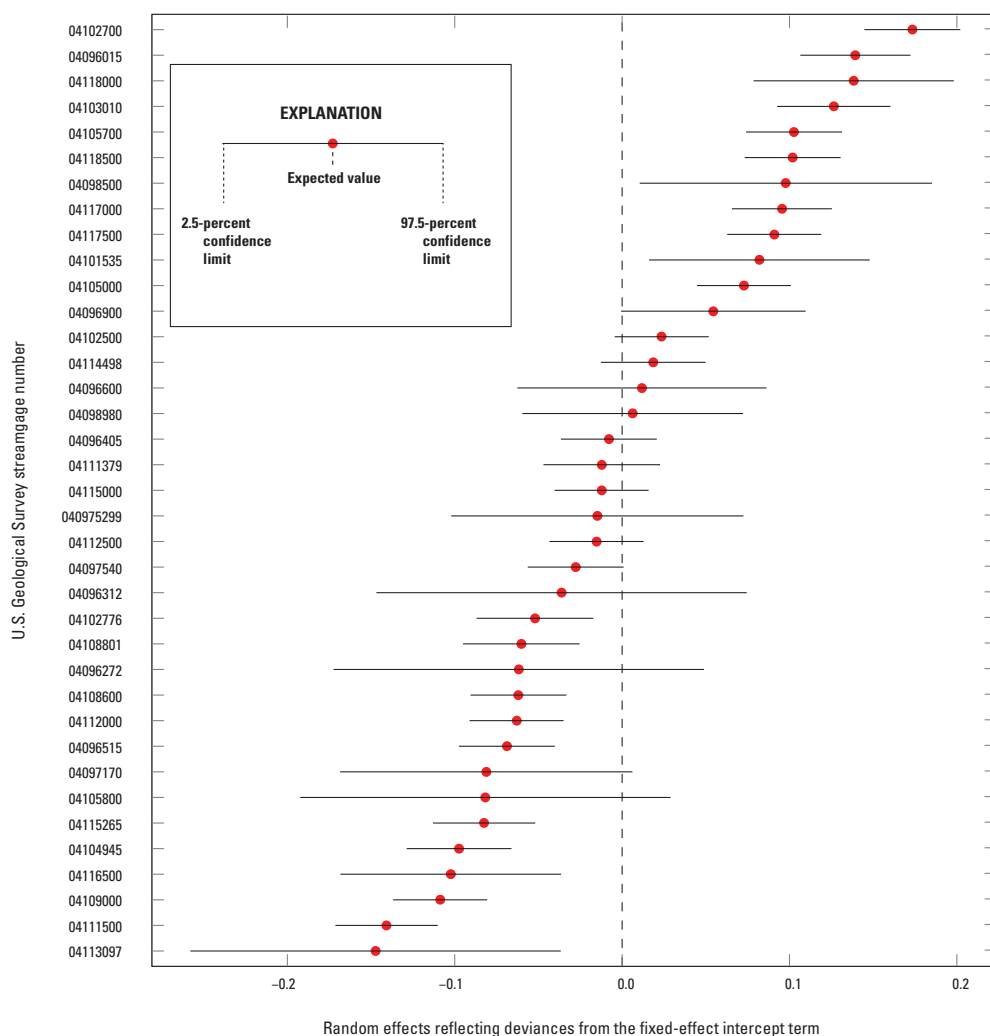


Prepared in cooperation with the Michigan Department of Environmental Quality

Basin, Climatic, and Irrigation Factors Associated with Median Summer Water Yields for Streams in Southwestern Michigan, 1945–2015



Scientific Investigations Report 2018–5071

Cover figure. Random effects reflecting deviances from the fixed-effect intercept term with 95-percent confidence intervals.

Basin, Climatic, and Irrigation Factors Associated with Median Summer Water Yields for Streams in Southwestern Michigan, 1945–2015

By David J. Holtschlag

Prepared in cooperation with the Michigan Department of Environmental Quality

Scientific Investigations Report 2018–5071

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

U.S. Department of the Interior
DAVID BERNHARDT, Acting Secretary

U.S. Geological Survey
James F. Reilly II, Director

U.S. Geological Survey, Reston, Virginia: 2019

For more information on the USGS—the Federal source for science about the Earth, its natural and living resources, natural hazards, and the environment—visit <http://www.usgs.gov> or call 1–888–ASK–USGS.

For an overview of USGS information products, including maps, imagery, and publications, visit <http://www.usgs.gov/pubprod/>.

Any use of trade, firm, or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Although this information product, for the most part, is in the public domain, it also may contain copyrighted materials as noted in the text. Permission to reproduce copyrighted items must be secured from the copyright owner.

Suggested citation:

Holtschlag, D.J., 2019, Basin, climatic, and irrigation factors associated with median summer water yields for streams in southwestern Michigan, 1945–2015: U.S. Geological Survey Scientific Investigations Report 2018–5071, 23 p., <https://doi.org/10.3133/sir20185071>.

ISSN 2328-0328 (online)

Acknowledgments

The author acknowledges the contribution of Andrew Lebaron, Michigan Department of Environmental Quality, who provided essential data on annual county irrigation in Michigan.

U.S. Geological Survey colleagues, David L. Lorenz (retired) and Casey J. Lee, provided numerous comments and suggestions that improved the quality of the report as part of their colleague reviews.

Contents

Aknowledgments	iii
Abstract	1
Introduction.....	1
Purpose and Scope	1
Previous Investigations.....	2
Description of Study Area	2
Annual Series of Median Summer Flows and the Period of Record Index Flow Statistic	2
Flows, Water Yields, and Transforms	2
Selection of Streamgages.....	6
Distribution of Median Summer Water Yields and Index Water Yields	6
Trends in Water Yields at Active Streamgages by Year	6
Estimation of Median Summer Water Yields Using Mixed Models.....	9
Form of the Linear Mixed Regression Models	9
Explanatory Variables	9
Time-Invariant Basin Characteristics.....	9
National Land Cover Database of 2011	10
State Soil Geographic Database	10
Hydrogeologic Properties	10
Time-Varying Climatic and Irrigation Data.....	10
Monthly Average Precipitation Data	10
Monthly Average Temperature Data	11
Annual Irrigation Data.....	12
Estimation of the Mixed Models.....	13
San Irrigation Full Model, 1945–2015.....	13
San Irrigation Step-Reduced Model, 1945–2015	13
San Irrigation Step-Reduced Model, 1970–2015	14
Irrigation Step-Reduced Model, 1970–2015	15
Estimation of Median Summer Water Yields at Streamgages in Southwestern Michigan.....	18
Summary.....	20
References Cited.....	21

Figures

1. Map showing the 40 selected U.S. Geological Survey streamgages in southwestern Michigan for analysis of water yields3
2. Boxplots showing the distribution of median summer water yields with index water yields for selected streamgages used in this analysis and in Hamilton and others (2008) in southwestern Michigan, 1945–20157
3. Graph showing the time series of annual median of median summer water yields for active streamgages in southwestern Michigan, 1945–2015.....8
4. Graph showing the time series of median summer precipitation in southwestern Michigan, 1945–201511

Figures—Continued

5. Jitter plot showing the time series of available annual irrigation by selected streamgages in southwestern Michigan, 1970–2015.....	12
6. Graph showing random effects reflecting deviances from the fixed-effect intercept term with 95-percent confidence intervals.....	16
7. Graph showing the relation between measured and predicted median summer water yields based on the Irrigation Step-Reduced Model, 1970–2015, for selected streamgages in southwestern Michigan.....	17
8. Screen capture showing R script to compute the magnitude and prediction intervals for the square root of median summer water yields in southwestern Michigan streams using mixed model object and specified data	19
9. Graph showing the distribution of estimated median summer water yields for U.S. Geological Survey streamgage 04102700 and other stream sites with the same basin and climatic characteristics as those at 04102700.....	20

Tables

1. Data for the 40 selected U.S. Geological Survey streamgages in southwestern Michigan.....	4
2. Weighted linear regression estimate of trend in median summer water yields in southwestern Michigan, 1945–2015.....	8
3. Variables and parameters in the fixed-effect component of the San Irrigation Step-Reduced Model, 1945–2015, for estimating the square roots of median summer water yields in southwestern Michigan.....	14
4. Variables and parameters in the fixed-effect component of the San Irrigation Step-Reduced Model, 1970–2015, for estimating the square roots of median summer water yields in southwestern Michigan.....	15
5. Variables and parameters in the fixed-effect component of the Irrigation Step-Reduced Model, 1970–2015, for estimating the square roots of median summer water yields in southwestern Michigan.....	15

Conversion Factors

U.S. customary units to International System of Units

Multiply	By	To obtain
Length		
inch (in.)	25.4	millimeter (mm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
acre	0.004047	square kilometer (km ²)
square mile (mi ²)	2.590	square kilometer (km ²)
Volume		
gallon (gal)	3.785	liter (L)
gallon (gal)	0.003785	cubic meter (m ³)
million gallons (Mgal)	3,785	cubic meter (m ³)
Flow rate		
cubic foot per second (ft ³ /s)	0.02832	cubic meter per second (m ³ /s)
cubic foot per second per square mile [(ft ³ /s)/mi ²]	0.01093	cubic meter per second per square kilometer [(m ³ /s)/km ²]
gallon per day (gal/d)	0.003785	cubic meter per day (m ³ /d)
million gallons per day (Mgal/d)	0.04381	cubic meter per second (m ³ /s)
million gallons per day per square mile [(Mgal/d)/mi ²]	1,461	cubic meter per day per square kilometer [(m ³ /d)/km ²]
inch per year (in/yr)	25.4	millimeter per year (mm/yr)
Hydraulic conductivity		
foot per day (ft/d)	0.3048	meter per day (m/d)
Transmissivity*		
foot squared per day (ft ² /d)	0.09290	meter squared per day (m ² /d)

*Transmissivity: The standard unit for transmissivity is cubic foot per day per square foot times foot of aquifer thickness [(ft³/d)/ft²]ft. In this report, the mathematically reduced form, foot squared per day (ft²/d), is used for convenience.

Temperature in degrees Celsius (°C) may be converted to degrees Fahrenheit (°F) as follows:

$$^{\circ}\text{F} = (1.8 \times ^{\circ}\text{C}) + 32.$$

Abbreviations

<i>BarrenP</i>	percentage of barren land
COMID	unique identifier for flowlines (stream segments) in NHDPlus
<i>cultivatedP</i>	percentage of land corresponding to cultivated crops
<i>cyear</i>	calendar year
<i>DevelopedP</i>	percentages of land in the developed open space, and low-, medium-, and high-intensity development spaces
<i>ForestP</i>	sum of percentages for deciduous forest, evergreen forest, and mixed forest
<i>HerbaceousP</i>	sum of percentages for grassland herbaceous, sedge herbaceous, lichens, and moss
<i>K1</i>	provides a measure of the average horizontal hydraulic conductivity on a scale ranging from 0 to 100 feet per day for the top 70 feet of unconsolidated deposits below land surface
<i>K2</i>	provides a measure of the average horizontal hydraulic conductivity of the full depth of unconsolidated deposits on a scale ranging from 0 to 100 feet per day.
MDEQ	Michigan Department of Environmental Quality
<i>Mean_TMP_degC</i>	Mean summer air temperature, in degrees Celsius
NHDPlus	National Hydrologic Dataset Plus value added attributes
NLCD 2011	National Land Cover Database of 2011
PDF	Adobe® portable document format
<i>p-value</i>	the probability that the null hypothesis of no statistical significance is true given the data
R	An open source programming language and software environment for statistical computing and graphics.
<i>R²</i>	coefficient of determination
<i>r_irrigMGDsqmi</i>	square root of annual irrigation estimates, in million gallons per day per square mile
<i>r_ppt_jun</i>	square root of annual June precipitation, in inches
<i>r_ppt_sum</i>	square root of the annual average monthly summer precipitation during July, August, and September
<i>ShrubP</i>	percentage of area in dwarf scrub or shrub scrub
<i>site</i>	intercept
<i>soil_A</i>	soil textures range from sand to sandy loams with high infiltration rates
<i>soil_B</i>	soil texture range from silt loam to loams with moderate infiltration rates
<i>soil_C</i>	soils textures are sandy clay loams with low infiltration rates
<i>soil_D</i>	soil textures range from clay loam to clay with very low infiltration rates
STATSGO	State Soil Geographic database
<i>T2</i>	texture-based estimated equivalent transmissivity of the glacial deposits, in square feet per day
USGS	U.S. Geological Survey
<i>WaterP</i>	percentage of the basin with open water
<i>WetlandP</i>	sum of percentages for woody wetland and emergent herbaceous wetlands
<i>zyear</i>	the calendar year minus 1945

Basin, Climatic, and Irrigation Factors Associated with Median Summer Water Yields for Streams in Southwestern Michigan, 1945–2015

By David J. Holtschlag

Abstract

Median summer water yields and resultant flows for streams are used in Michigan to regulate large water withdrawals to help prevent negative effects on characteristic fish populations. Large water withdrawals commonly are associated with irrigation in rural areas. In an earlier statewide report, an index-flow statistic for the period of record, defined as the median flow during the summer month of lowest flow, was used to characterize median summer flows and associated water yields. In this report, the annual series of median summer water yields for the period July 1 through September 30 within the period of record is used to characterize median summer water yields. For 27 streamgages included in both reports, the average index water yield was at the 37th percentile of the distribution of median summer water yields. In contrast to an index statistic, an annual time series provides a basis for detecting trends in median summer water yields and for determining basin, climatic, and irrigation factors affecting spatial and temporal variations in summer water yields. Daily flow data from 40 selected U.S. Geological Survey streamgages in southwestern Michigan were used in this analysis. Two mixed models were identified to estimate median summer water yields based on fixed basin characteristics and temporally varying climatic factors for 1945–2015. No irrigation data were available prior to 1970, so no irrigation variables were included in the mixed models for 1945–2015. Then, two mixed models were developed for 1970–2015, a period in which a partial annual series of county-level irrigation data also were available. One of the 1970–2015 mixed models provides a basis for estimating median summer water yields at sites in southwestern Michigan using an estimated trend component, and selected basin, climatic, and irrigation factors. Re-estimation of model parameters in this mixed model with more spatially precise information on irrigation withdrawals may improve model accuracy.

Introduction

The State of Michigan (2006) passed Public Act 33 (PA33–2006) to regulate water withdrawals. The legislation sought to prevent any new or increased large-capacity withdrawal, defined as a withdrawal that averages more than 100,000 gallons of water per day (0.1547 cubic foot per second) in any consecutive 30-day period, from adversely affecting resources. Resources are considered adversely affected if the flow of a stream is decreased by part of the index flow such that the stream's ability to support characteristic fish populations is functionally impaired. Within the legislation, the index flow was defined as the 50-percent exceedance (median) flow for the lowest flow month of the flow regime in the period of record. Large water withdrawals commonly are associated with irrigation in rural areas. In the statewide application by Hamilton and others (2008), the index flow was specified as the median flow during the summer month of lowest flow in Michigan, which varied among selected streamgages.

Effective implementation of this legislation may benefit from partitioning the effects of spatial and temporal factors on the summer median water yields. This partitioning may help differentiate natural and manageable factors associated with changes in summer median water yields. The index flow equation (Hamilton and others, 2008, p. 28) does not provide a basis for characterizing temporal effects because none of the explanatory variables are time varying.

Purpose and Scope

This report develops an alternative to a period-of-record index water yield statistic used in a statewide study by Hamilton and others (2008) for characterizing median water yield during the summer month of lowest median water yield in Michigan. The alternative is based on an annual time series of

median daily water yields during the 92-day period from July 1 through September 30 of each year, which is referred to as the annual series of median summer water yields in this report. This time series provides a basis for characterizing the time-invariant effects of basin characteristics and the time-varying effects of climatic factors and irrigation withdrawals on median summer water yields. A mixed model is developed to assess the effectiveness of this alternative by relating repeated measures of median summer water yields to basin, irrigation, and climatic characteristics at 40 selected U.S. Geological Survey (USGS) streamgages in southwestern Michigan. Application of a mixed model is described for sites with similar characteristics.

Previous Investigations

In response to PA33–2006 by the State of Michigan (2006), Hamilton and others (2008) developed a regression model for computing an index flow describing the median flow for the summer month of lowest flow in Michigan; summer months were July, August, and September. This single median flow statistic represented the period of record, rather than an annual statistic. To facilitate regionalization, the index flow was divided by the corresponding drainage area of the basin at the monitoring streamgage to compute an index water yield. The square root transformation of water yield was used as the hydrologic response in the statewide analysis (Hamilton and others, 2008). Squaring the estimated response and multiplying by the drainage area at the site of interest would provide an index flow estimate.

The multiple regression equation developed by Hamilton and others (2008, p. 28) included the percentage of basin underlain by low-transmissivity (*L_Trans*) and high-transmissivity (*H_Trans*) aquifer materials, percentage of forest area (*Forest*), normal (1971–2000) annual precipitation (*Precip*), and percentage of basin overlain with group *A* (*A_Soils*) and group *D* (*D_Soils*) hydrologic soils. The regression estimates explained about 72 percent of the variation in square roots of water yields at 147 streamgages included in the analysis.

Description of Study Area

Michigan is in the eastern north-central part of the United States and consists of two peninsulas separated by the Straits of Mackinac, a body of water that connects Lake Michigan with Lake Huron. In this report, southwestern Michigan refers to an area generally corresponding to a multicounty area spanning hydrologic unit code 0405 (Seaber and others, 1994) in the Lower Peninsula (fig. 1).

Southwestern Michigan has a humid continental climate in which the average precipitation varies from about 33 to 38 inches per year (fig. 10 on p. 16 of Hamilton and others, 2008). Southwestern Michigan is bordered on the west by Lake Michigan, where temperature and precipitation patterns may be locally affected. Average snowfall depths generally range from 50 to 70 inches per year (fig. 11 on p. 17 of Hamilton and

others, 2008). Hydrologic soil groups B and C are more common in southwestern Michigan than soil groups A and D (fig. 9 on p. 15 of Hamilton and others, 2008). Forests generally cover less than 40 percent of southwestern Michigan (fig. 8 on p. 14 of Hamilton and others, 2008). The spatial variation of hydraulic conductivity and transmissivity is documented by Bayless and others (2017). Water withdrawals associated with irrigation in Michigan are compiled and aggregated to the county level by the Michigan Department of Environmental Quality (MDEQ; Michigan Department of Environmental Quality, 2018; Michigan Department of Environmental Quality and others, 2018).

Annual Series of Median Summer Flows and the Period of Record Index Flow Statistic

Hamilton and others (2008) determined that the summer month of lowest median flow used to compute the index flow varied among the 147 streamgages used in the statewide analysis. At 5 streamgages, the lowest median flow was in July, at 92 streamgages the lowest median flow was in August, and at 50 streamgages the lowest median flow was in September. In addition, although a single summer month may have the lowest median flow during the period of record, the other summer months commonly have a lower median flow during individual years.

Flows, Water Yields, and Transforms

Consistent with Hamilton and others (2008), graphs and analysis in this report are generally presented on the basis of water yields rather than flows. Water yields are computed as flows divided by the drainage areas of the corresponding basin monitored by a streamgage. Drainage areas are based on topographic features forming surface-water divides, which are thought to be generally consistent with groundwater divides. This transformation facilitates comparison of flow characteristics from basins that vary widely in drainage areas. For example, USGS streamgage 04117000 (Quaker Brook near Nashville, Michigan) monitors flow from a basin with a drainage area of 7.6 square miles, and streamgage 04118000 (Thornapple River near Caledonia, Mich.) has a drainage area of 773 square miles (table 1). So, although the two index flows of 2.80 and 281 cubic feet per second differ by two orders of magnitude, corresponding water yields are similar at 0.359 and 0.353 cubic foot per second per square mile, respectively.

Water yields, like flows, tend to be skewed to the right and bounded by zero on the left. A square-root transformation was applied so that the distribution of water yields would more closely approximate a normal distribution but still accommodate zero values. Squaring model estimates and multiplying by the drainage area at the site of interest would provide estimates in cubic feet per second.

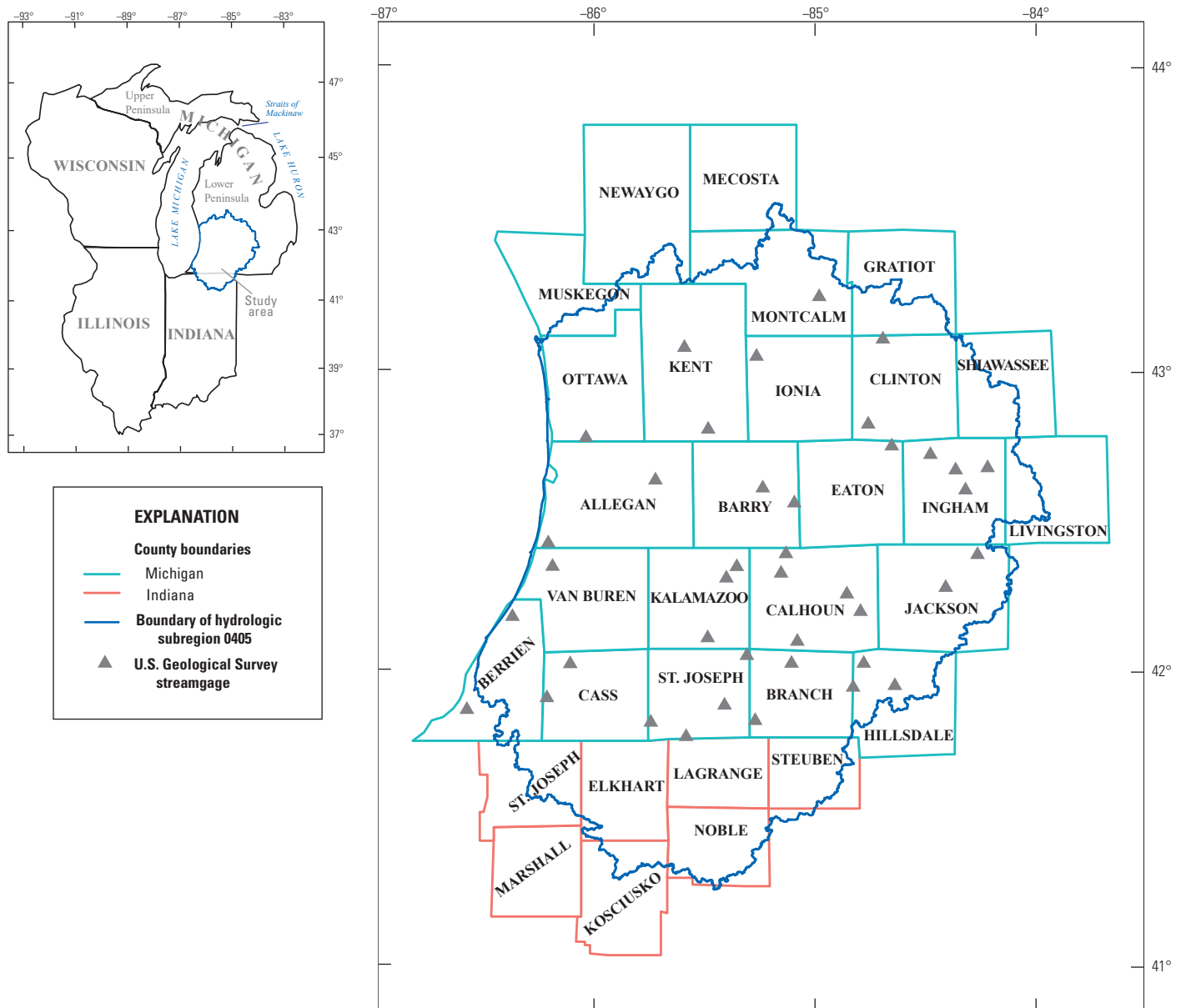


Figure 1. The 40 selected U.S. Geological Survey streamgages in southwestern Michigan for analysis of water yields.

Table 1. Data for the 40 selected U.S. Geological Survey streamgages in southwestern Michigan.

[MI, Michigan; W, West; Hwy, Highway]

U.S. Geological Survey streamgage number	U.S. Geological Survey streamgage name	Decimal latitude of streamgage	Decimal longitude of streamgage	Basin area (square miles)	Beginning year of streamflow record used in analysis	Ending year of streamflow record used in analysis	Years of record used in analysis
04096015	Galien River near Sawyer, MI	41.8737	−86.5750	80.7	1995	2015	19
04096272	Beebe Creek at Moore Road near Hillsdale, MI	41.9542	−84.6388	42.4	1974	1980	7
04096312	Sand Creek at Litchfield, MI	42.0292	−84.7797	20.6	1974	1977	4
04096405	St. Joseph River at Burlington, MI	42.1031	−85.0800	206	1963	2015	53
04096515	South Branch Hog Creek near Allen, MI	41.9487	−84.8277	48.7	1970	2015	46
04096600	Coldwater River near Hodunk, MI	42.0292	−85.1064	293	1963	1989	27
04096900	Nottawa Creek near Athens, MI	42.0553	−85.3084	162	1967	1997	31
04097170	Portage River at W Avenue near Vicksburg, MI	42.1148	−85.4856	68.2	1946	1979	21
040975299	Prairie River at US 12 near Burr Oak, MI	41.8364	−85.2708	66	2014	2015	2
04097540	Prairie River near Nottawa, MI	41.8884	−85.4094	106	1963	2015	53
04098500	Fawn River near White Pigeon, MI	41.7823	−85.5833	192	1958	1975	18
04098980	Mill Creek at Quaker Street near Mottville, MI	41.8306	−85.7430	43.5	2012	2015	4
04101535	Dowagiac River at State Hwy 51 near Dowagiac, MI	42.0278	−86.1075	74.3	2012	2015	4
04101800	Dowagiac River at Sumnerville, MI	41.9134	−86.2131	255	1961	2015	55
04102500	Paw Paw River at Riverside, MI	42.1864	−86.3689	390	1952	2015	64
04102700	South Branch Black River near Bangor, MI	42.3542	−86.1875	83.6	1966	2015	50
04102776	Middle Branch Black River near South Haven, MI	42.4325	−86.2070	83	1995	2011	17
04102850	South Branch Kalamazoo River near Albion, MI	42.2031	−84.7941	146	1972	1976	5
04103010	Kalamazoo River near Marengo, MI	42.2617	−84.8558	267	1987	2011	25
04104945	Wanadoga Creek near Battle Creek, MI	42.3964	−85.1317	48.3	1995	2015	21
04105000	Battle Creek at Battle Creek, MI	42.3314	−85.1536	241	1945	2015	70
04105700	Augusta Creek near Augusta, MI	42.3534	−85.3539	38.9	1965	2015	51
04105800	Gull Creek at 37th Street near Galesburg, MI	42.3150	−85.4011	38.1	1965	1972	8
04108600	Rabbit River near Hopkins, MI	42.6423	−85.7220	71.4	1966	2015	50
04108801	Macatawa River near Zeeland, MI	42.7836	−86.0364	68.5	1961	2007	47
04109000	Grand River at Jackson, MI	42.2836	−84.4088	174	1945	2015	70
04110000	Orchard Creek at State Hwy-106 near Munith, MI	42.3936	−84.2650	49	1945	1956	11

Table 1. Data for the 40 selected U.S. Geological Survey streamgages in southwestern Michigan.—Continued

[MI, Michigan; W, West; Hwy, Highway]

U.S. Geological Survey streamgage number	U.S. Geological Survey streamgage name	Decimal latitude of streamgage	Decimal longitude of streamgage	Basin area (square miles)	Beginning year of streamflow record used in analysis	Ending year of streamflow record used in analysis	Years of record used in analysis
04111379	Red Cedar River near Williamston, MI	42.6831	−84.2191	163	1975	2015	29
04111500	Deer Creek near Dansville, MI	42.6092	−84.3191	16.3	1954	2011	58
04112000	Sloan Creek near Williamston, MI	42.6759	−84.3639	9.34	1954	2015	62
04112500	Red Cedar River at East Lansing, MI	42.7273	−84.4780	355	1945	2015	70
04113097	Carrier Creek at Willow Hwy near Lansing, MI	42.7556	−84.6528	12.1	1975	1980	6
04114498	Looking Glass River near Eagle, MI	42.8281	−84.7594	280	1945	2015	65
04115000	Maple River at Maple Rapids, MI	43.1098	−84.6931	434	1945	2015	70
04115265	Fish Creek near Crystal, MI	43.2498	−84.9811	39.7	1988	2015	28
04116500	Flat River at Smyrna, MI	43.0528	−85.2647	528	1951	1986	36
04117000	Quaker Brook near Nashville, MI	42.5659	−85.0936	7.6	1954	2015	43
04117500	Thornapple River near Hastings, MI	42.6159	−85.2364	385	1945	2015	70
04118000	Thornapple River near Caledonia, MI	42.8111	−85.4834	773	1952	1994	42
04118500	Rogue River near Rockford, MI	43.0822	−85.5909	234	1952	2015	59

Selection of Streamgages

USGS streamgages in southwestern Michigan were selected to represent mainly rural areas where natural flow conditions may be affected by surface or groundwater withdrawals for irrigation. To ensure that selected streamgages were sensitive to local withdrawals, the maximum drainage area of a gaged basin was restricted to less than 1,000 square miles. Given the utility of even a few summer median flows, a minimum of 2 years of continuous flow data were needed at streamgage for inclusion in the study. Streams were not selected if summer flows might be affected by regulation from flow retention or by natural storage in large lakes. As a result, data from 40 streamgages were included in this analysis (fig. 1; table 1); streamgage data are available from the USGS National Water Information System database (U.S. Geological Survey, 2017).

The number of active streamgages within this selected set varied during the period of analysis from 1945 to 2015, when ancillary data on monthly precipitation and temperature were available. A minimum of 7 streamgages were operating in 1945, which steadily increased to a maximum of 30 streamgages in 1975, and then decreased to 24 in 2015. Streamgage 040975299 (Prairie River at US 12 near Burr Oak, Mich.) had the shortest period of record with 2 years of data, and five streamgages were active during the entire period of analysis with 71 years of data.

Distribution of Median Summer Water Yields and Index Water Yields

Of the 40 streamgages in southwestern Michigan selected for this analysis, 27 also were included in the statewide report by Hamilton and others (2008). A comparison of the distribution of median summer water yields and index water yields is shown in figure 2. At 22 streamgages, the index water yield was between the 25th and 50th percentile of median of the median summer water yield, while the index flow was below the 25th percentile at 3 streamgages. At 2 streamgages, the index flow was between the 50th and 75th percentile of the distribution of median summer water yields. The index yield, corresponding to the long-term median (50th percentile) yield during the summer month of lowest flow, corresponds to the 37th percentile over the 27 streamgages in the Hamilton and others (2008) report. Thus, the 50th percentile of the median summer water yields is somewhat higher than the index water yields.

Trends in Water Yields at Active Streamgages by Year

A possible difficulty with the index water yield statistic (Hamilton and others, 2008) is that time series of yields, taken annually over all active streamgages, may have a trend. If a trend is present, then the value of the index statistic would only represent the time interval over which the statistic was computed rather than characterizing an invariant property of streamflow. Hodgkins and others (2007) indicate increasing water yields (runoff) in southwestern Michigan in July, August, and September and generally increasing precipitation in August and September.

To assess possible trends in the median summer water yields in southwestern Michigan from 1945 to 2016, the annual median of median summer water yields was computed for each year of record based on data from active streamgages (fig. 3). A weighted linear regression was used to estimate the relation between the square root of median summer water yields ($\sqrt{\text{med_ylds_yr}}$) with the calendar year (cyear) and the previous calendar year's (lag_1) square root of median summer water yields ($\text{lag_sqrt_med_ylds_yr}$), where weights were the number of streamgages operating in that year (n_sites_yr). The results indicate that there is a positive trend in the square root of median summer water yields in southwestern Michigan (the probability that the null hypothesis of no statistical significance is true given the data [p -value]=0.0142). The lag_1 term was included to reduce the possibility of inflating the significance of the trend test because of autocorrelation in annual series of median summer water yields. The trend line shown in figure 3 is based on the slope parameter estimate of 0.001619 for cyear (table 2) fitted through the median year and median water yield for the period of analysis. The residual standard error is 0.4193 on 68 degrees of freedom. The adjusted multiple coefficient of determination (R^2) is 0.2186.

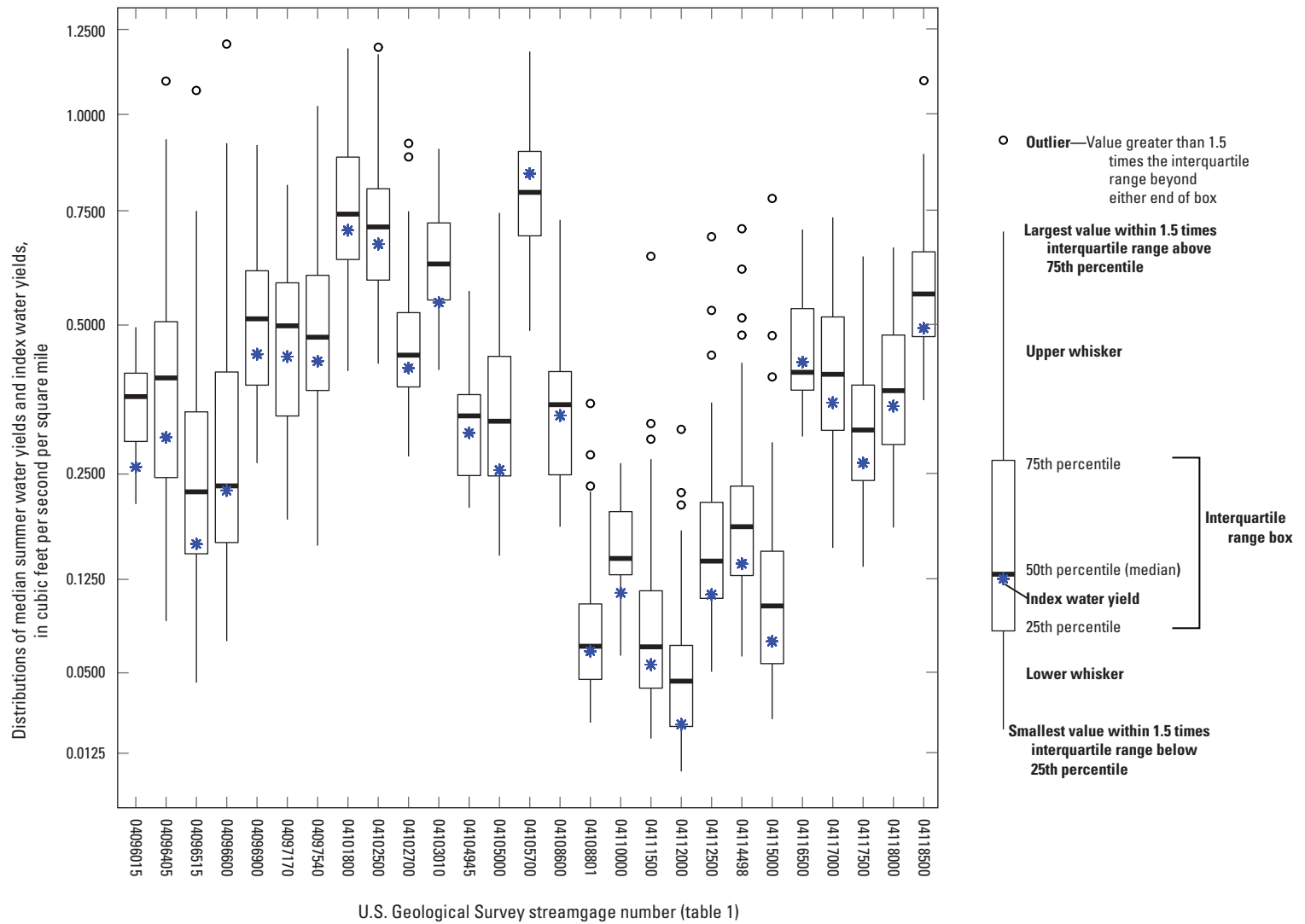


Figure 2. The distribution of median summer water yields with index water yields for selected streamgages used in this analysis and in Hamilton and others (2008) in southwestern Michigan, 1945–2015.

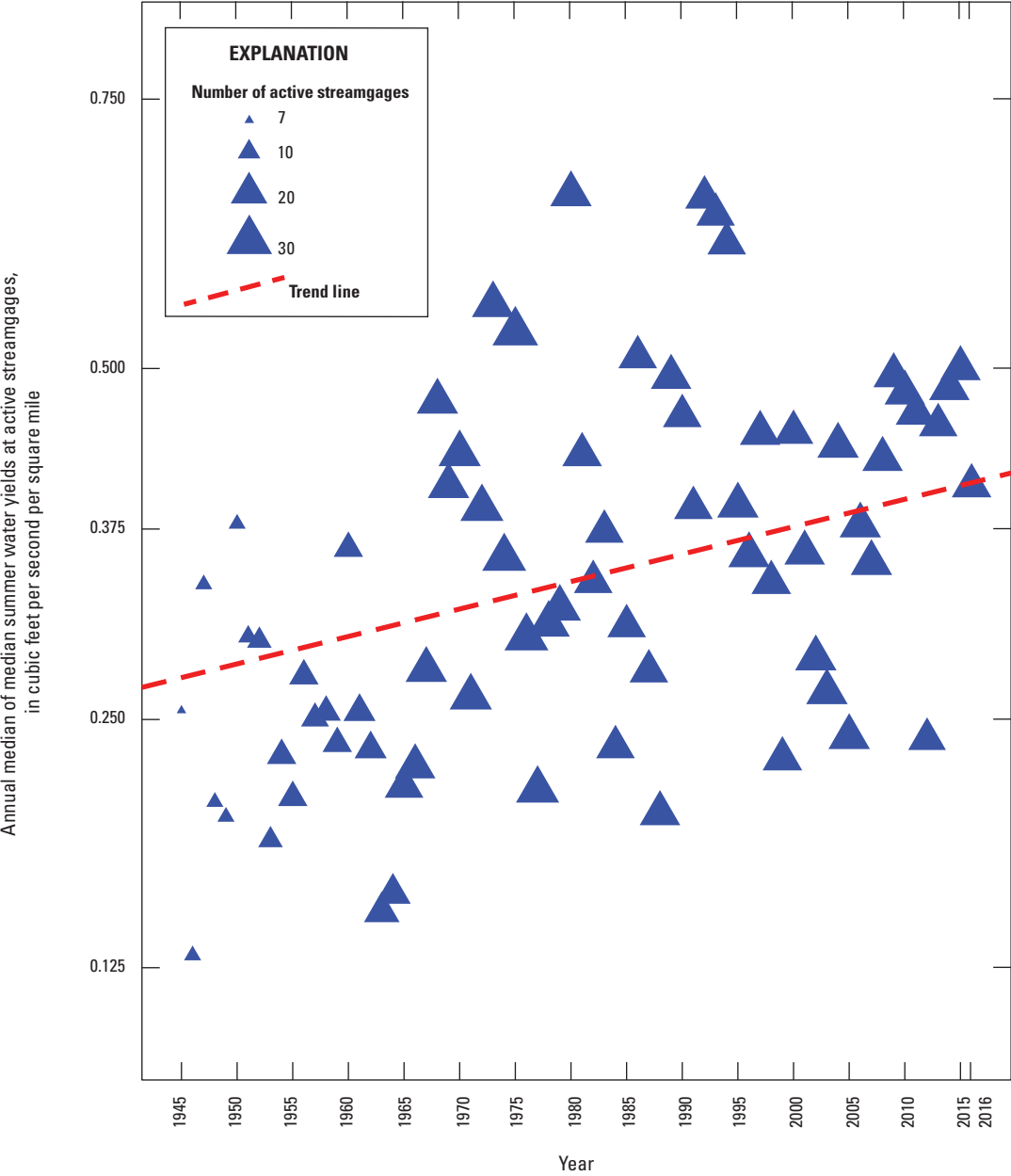


Figure 3. The annual median of median summer water yields for active streamgages in southwestern Michigan, 1945–2015.

Table 2. Weighted linear regression estimate of trend in median summer water yields in southwestern Michigan, 1945–2015.

[*t*, a statistic formed as the ratio of the difference between estimated and expected values divided by its standard error; *t*-value, parameter estimate divided by the standard error; Prob, probability; >, greater than; *cyear*, calendar year; *lag_sqrt_med_ylds_yr*, previous year's (*lag*) square root of median water yields]

Variable	Parameter estimate	Parameter standard error	<i>t</i> -value	Prob(> <i>t</i>)
Intercept	−2.7870	1.2487	−2.232	0.0289
<i>cyear</i>	0.001619	0.0006431	2.518	0.0142
<i>lag_sqrt_med_ylds_yr</i>	0.2847	0.1157	2.462	0.0164

Estimation of Median Summer Water Yields Using Mixed Models

Linear mixed (regression) models were developed to estimate median summer water yields. Mixed models were used because these models accommodate repeated measurements on multiple sites without violating model error assumptions implicit in ordinary least squares regression (Draper and Smith, 1998). Through a mixed model, annual estimates of median summer water yields can be conditioned on presumably fixed basin characteristics, and time-varying climatic variables and irrigation rates. The discussion that follows presents the form of the mixed model and a discussion of the explanatory variables used in the analysis.

Form of the Linear Mixed Regression Models

The linear mixed regression model, or mixed model, for a site (i) and year (j), can be represented as:

$$y_{ij} = X \times \beta + Z \times \gamma + \varepsilon_{ij} \quad (1)$$

where

- y_{ij} is an $I \times J$ column vector of the square root of median summer water yields for all $i \in I$ streamgages used in the report and $j \in J$ years in the analysis,
- X is an $I \times J$ by $\#P$ design matrix where $\#P$ is the number of fixed effects,
- β is a $\#P$ vector of estimated fixed-effect parameters,
- Z is an $I \times J$ by $\#Q$ design matrix of random effects where $\#Q$ is the number of included sites,
- γ is a $\#Q$ vector of estimated random intercept deviations from the fixed intercept, and
- ε_{ij} is an $I \times J$ vector of random errors for all sites (i) and water yields for all years (j).

The dimension of the Z matrix presumes that the random component consists only of different intercept terms for each site.

Both fixed and random effects are included in a mixed model. Fixed effects describe the average statistical relation between explanatory variables and water yields for all basins in the study area, much like an ordinary least squares regression model. Random effects, on the other hand, describe deviances from these average statistical relations that are applicable to a site. Random effects can be associated with the average response, when applied to an intercept term, or to individual explanatory variables, when the corresponding parameters are viewed as slope terms.

In this report, a random intercept term associated with the streamgage number was identified to describe deviances from the intercept term in the fixed component. The intercept deviances in the random component are constrained to sum to zero. The standard deviation of the intercept deviances reflects the expected variability of these deviances in the population of sites sampled by the 40 streamgages included in the study. No other random effects were included in the mixed model.

Estimated confidence intervals for parameters in the fixed component are based on a Wald approximation (Bates and others, 2015), and their statistical significances were evaluated based on a Student's t -test (Iman and Conover, 1983). Correspondingly, small probabilities for t -statistic indicate that the null hypothesis that parameters are zero can be rejected. The significance of the random component was tested by comparing the likelihood ratio test statistic with the chi-square distribution (Bates and others, 2015).

Explanatory Variables

In this report, explanatory variables correspond to time invariant basin characteristics, such as the aquifer or soil properties, and time-varying features, such as precipitation, temperature, or irrigation, that may be reflected in water yields. Basin and selected climatic characteristics used in this report were generally derived from the National Hydrologic Dataset Plus value added attributes (NHDPlus, version 2) medium resolution (1:100,000 scale) coverages (McKay and others, 2017). NHDPlus catchments were matched to catchments of streamgage basins by the unique identifier for flowlines (stream segments) in NHDPlus (COMID). NHDPlus catchments average about 1 square mile, and when all catchments within a given drainage basin are summed, they generally provide a close approximation to the gaged area of the basin.

Time-Invariant Basin Characteristics

In this report, time-invariant basin characteristics refer to characteristics that are generally thought to vary slowly with time when considered over extensive areas. Given the rural nature of southwestern Michigan, attributes described in the National Land Cover Database of 2011 (NLCD 2011) and Homer and others (2015), for example, are thought to be indicative of the land cover conditions from 1945 to 2015. The percentage of basin with forest cover, which was included as an explanatory variable by Hamilton and others (2008), is among variables also defined in the NLCD 2011. Similarly, the State Soil Geographic (STASGO) database (U.S. Department of Agriculture, 1994) was used to determine the percentages of hydrologic soil groups, which were included in the index flow equation developed by Hamilton and others (2008, p. 25). The 1971–2000 normal precipitation included in Hamilton and others (2008) was replaced with selected monthly or seasonal climatic data to estimate median summer water yields.

National Land Cover Database of 2011

The National Land Cover Database of 2011 is available in NHDPlus (U.S. Environmental Protection Agency and U.S. Geological Survey, 2012). Similarly, metadata for the land cover data can be downloaded in file 0NHDPlusV2_NLCDLandUse_metadata_20141013.htm. The classification system used in NLCD 2011 is modified from the Anderson Land Cover Classification System (Anderson and others, 1976).

Land cover variables were used to develop explanatory variables for the regression analysis. The variable *WaterP* represents the percentage of the basin with open water (NLCD class 11). The variable *DevelopedP* represents the percentages of land in the developed open space (class 21), developed low intensity (class 22), developed medium intensity (class 23), and developed high intensity (class 24) classes. The variable *BarrenP* represents the percentage of barren land (class 31). The variable *ForestP* represents a major land use computed as the sum of percentages for deciduous forest (class 41), evergreen forest (class 42), and mixed forest (class 43). The variable *ShrubP* represents the percentage of area in dwarf scrub (class 51) or shrub scrub (class 52). The variable *HerbaceousP* is computed as the sum of percentages for grassland herbaceous (class 71), sedge herbaceous (class 72), lichens (class 73), and moss (class 74). The variable *cultivatedP* is the percentage of land corresponding only to cultivated crops (class 82). Total cultivated crops in NLCD 2011 includes both pasture (or hay) and cultivated crops (classes 81 and 82). The *WetlandP* variable is computed as the sum of percentages for woody wetland (class 90) and emergent herbaceous wetlands (class 95).

State Soil Geographic Database

Data for the STATSGO soil characteristics and hydrologic groups are available from ScienceBase (Wieczorek and others, 2016). The STATSGO_HYDGRP_CAT_CONUS.zip was downloaded, which provides catchment-based information on soil groups. Metadata describing individual fields can be accessed from the same page by downloading NHDV2_STATSGO_HYDGRP_CONUS.xml. Only homogeneous hydrologic groups A (variable *soil_A* [soil textures range from sand to sandy loams with high infiltration rates]), B (*soil_B* [soil textures range from silt loam to loams with moderate infiltration rates]), C (*soil_C* [soils textures are sandy clay loams with low infiltration rates]), and D (*soil_D* [soil textures range from clay loam to clay with very low infiltration rates]) were included as possible explanatory variables in the analysis. Mixed hydrologic soil groups, such as AC, AD, BC, BD, or CD, were not included.

Hydrogeologic Properties

Three hydrogeologic properties were included as possible explanatory variables in the regression analysis. The variables (*K1*, *K2*, and *T2*) are based on textural descriptions from water well driller logs obtained from State well log databases (Bayless and others, 2017). These variables are thought to provide hydrogeologic information that is comparable to the low and high transmissivity classes, which were statistically significant in Hamilton and others (2008). Weighted mean values for the three variables were computed by intersecting polygons of selected basins, formed from NHDPlus catchments, over georeferenced grids of hydrogeologic properties.

The variable *K1* provides a measure of the average horizontal hydraulic conductivity on a scale ranging from 0 to 100 feet per day for the top 70 feet of unconsolidated deposits below land surface. Similarly, the variable *K2* provides a measure of the average horizontal hydraulic conductivity of the full depth of unconsolidated deposits on a scale ranging from 0 to 100 feet per day. The *T2* variable represents the texture-based estimated equivalent transmissivity of the glacial deposits, in square feet per day (Bayless and others, 2017).

Time-Varying Climatic and Irrigation Data

Monthly climatic data, including precipitation and air temperature, and annual irrigation data were included in this analysis to help describe temporal variations in median summer water yields. The sources and characteristics of these data are described in the following paragraphs.

Monthly Average Precipitation Data

Monthly NHDPlus catchment-average estimates of precipitation for 1945 through 2015 are available from ScienceBase (Wieczorek and others, 2017a). Each file name downloaded is of the form PPT[yyyy]_ANN_CONUS.zip, where [yyyy] indicates the four-digit calendar year of data. The zipped folder for each year contains a single data file in American Standard Code for Information Interchange (ASCII) format. Each file contains monthly average precipitation data for about 2.7 million NHDPlus catchments in the conterminous United States. Each row of data within each file is uniquely indexed by an NHDPlus assigned COMID and followed by 12 monthly values of precipitation depth, in millimeters. Depths in millimeters were converted to inches.

The monthly data were used to form two variables for regression: spatially averaged precipitation over the basin boundary for June in inches (*ppt_jun_in*), and averaged for the summer months of July, August, and September (*ppt_sum_in*). The square root transformation was applied

to precipitation data to normalize their distribution, denoted with the prefix “r” to form *r_ppt_jun_in* and *r_ppt_sum_in*, respectively. A median value was computed for precipitation by year to describe the central tendency in basins with active streamgages.

A linear trend analysis of the annual median at active streamgages of average summer precipitation values was developed. A square root transformation was applied to summer precipitation values for the test, and annual series were weighted by the number of streamgages operated in the corresponding year. Results indicate that square roots of summer precipitation are likely increasing by 0.003173 square root inches per year (p -value=0.00745). A lag₁ autocorrelation term for transformed precipitation values was tested but determined to be insignificant (p -value=0.7628) and was not included in the trend model. Monthly precipitation values and the trend estimate are shown in figure 4.

Monthly Average Temperature Data

Monthly average temperature data from 1945 to 2015 are available from ScienceBase (Wieczorek and others, 2017b). Each file name is of the form TAV[yyyy]_CAT_CONUS.zip, where [yyyy] indicates the four-digit calendar year of data. The zipped folder for each year contains a single data file in ASCII format. Each file contains temperature data for approximately 2.7 million NHDPlus catchments in the conterminous United States. Each row of data within each file is uniquely indexed by COMID and followed by 12 monthly values of temperature, in degrees Celsius. The explanatory variable *mean_TMP_degC* (mean summer air temperature, in degrees Celsius) was formed for possible inclusion in the regression. Based on 71 years of data, the Mann-Kendall test (Iman and Conover, 1983) returned an *S* statistic of 65 with a corresponding p -value of 0.75. The null hypothesis of no trend in temperature was not rejected.

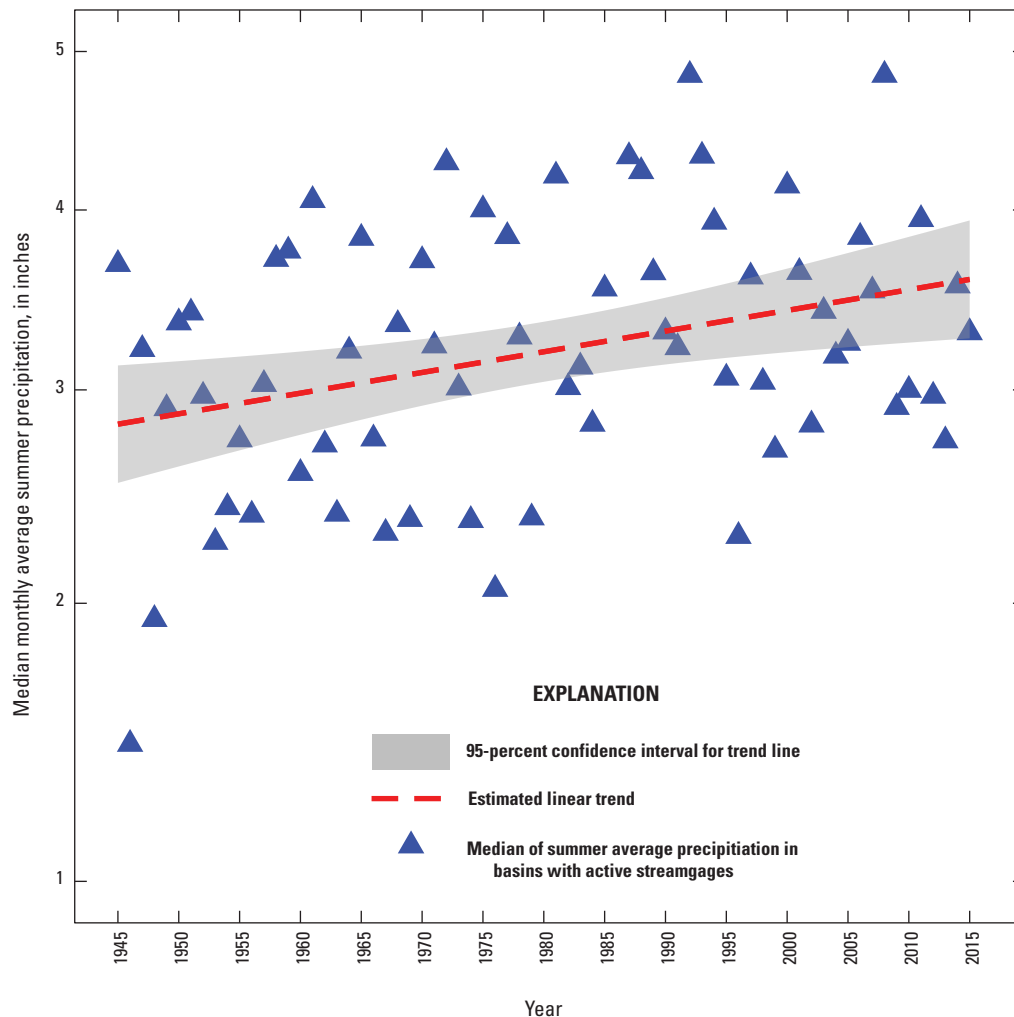


Figure 4. Time series of median summer precipitation in southwestern Michigan, 1945–2015.

Annual Irrigation Data

Unlike other time-varying attributes included as explanatory variables in this report, monthly irrigation data are not available on an NHDPlus catchment basis. Instead, annual irrigation data are aggregated to the county level (Michigan Department of Environmental Quality, 2018; Michigan Department of Environmental Quality and others, 2018). The MDEQ is the primary source of irrigation data for Michigan. The USGS, in cooperation with MDEQ, has prepared a series of 5-year reports on water use beginning in 1985 that are consistent with national standards (Hutson, 2007).

Documented irrigation is available for 24 years during 1945–2015: 1970, 1977, 1985, 1990, 1995, 1997–2015 (fig. 5). Bedell and Van Til (1979) provide some of the earliest statewide irrigation data for Michigan, including annual estimates for 1970 and 1977. USGS irrigation data, which was coordinated with MDEQ information, provided information

for 1985, 1990, and 1995 (Morenz and others, 2005). Annual irrigation data from 1997 through 2015 were obtained from reports published by MDEQ (Michigan Department of Environmental Quality, 2018; Michigan Department of Environmental Quality and others, 2018).

Annual irrigation data compiled by MDEQ and aggregated to the county level for 1997 to 2006 were obtained from Adobe® portable document format (PDF) files (Michigan Department of Environmental Quality and others, 2018). MDEQ annual irrigation data for 2007 to 2015, also aggregated to the county level, were obtained from Microsoft® Excel® files (Michigan Department of Environmental Quality, 2018).

The irrigation variable (*irrgMGDsqmi*) represented the depth of irrigation in million gallons per day per square mile of land area in the basin (fig. 5). The square root transformation was applied to help normalize their distribution, and the variable was referenced as *r_irrgMGDsqmi* (square root of annual irrigation estimates, in million gallons per day per square mile).

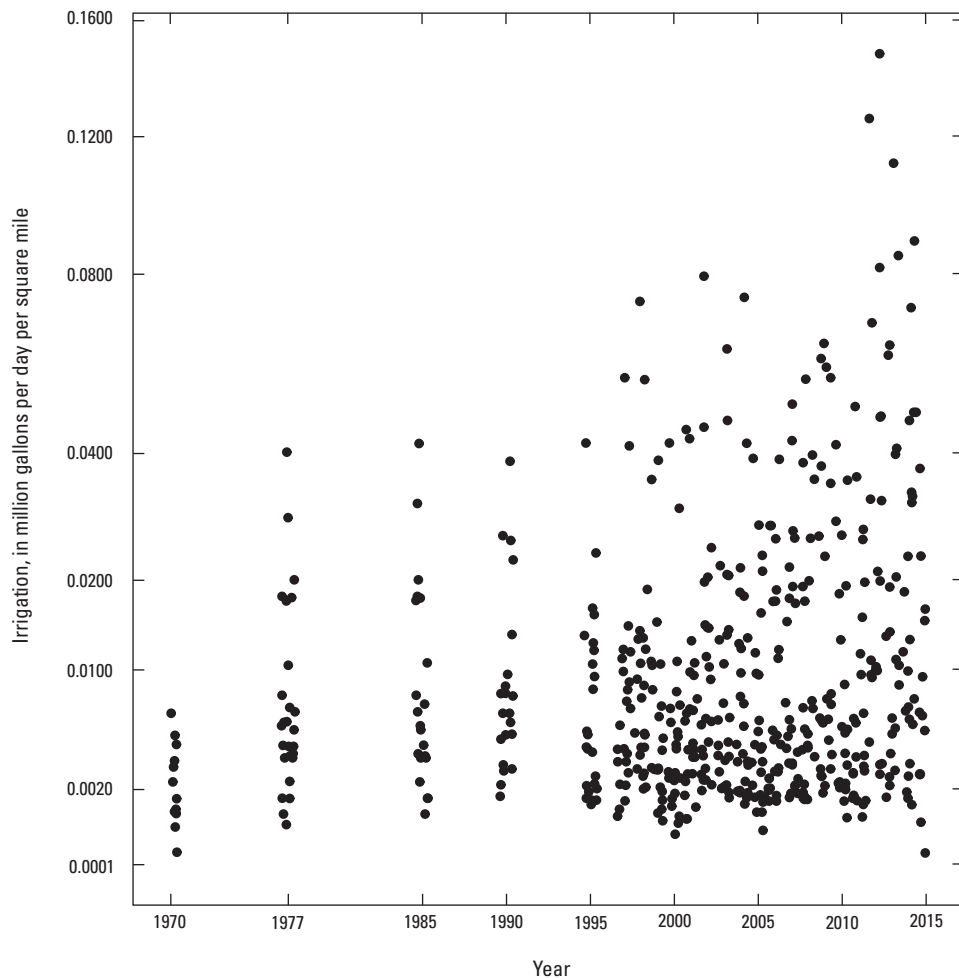


Figure 5. Time series of available annual irrigation by selected streamgages in southwestern Michigan, 1970–2015.

Estimation of the Mixed Models

Mixed models were estimated in the statistical programming language and software environment for statistical computing and graphics, R (R Core Team, 2017), using functions in packages lme4 (Bates and others, 2015) and lmerTest (Kuznetsova and others, 2016). The 40 selected streamgages described previously were used in the analysis during periods when supporting annual series of median summer water yields and explanatory features were available. The model response variable was median summer water yields with a square root transformation to help normalize the variate. Explanatory features were systematically evaluated for inclusion in the model.

Four mixed models were estimated. The San Irrigation Full Model, 1945–2015, includes data from 40 streamgages that were active for 2 or more years during the 71-year period from 1945 to 2015. This model includes 19 explanatory features and 1,478 observations of median summer flow, but no (san) irrigation data. The San Irrigation Step-Reduced Model, 1945–2015, contains only statistically significant variables identified by backward elimination from the San Irrigation Full Model, 1945–2015. The San Irrigation Step-Reduced Model, 1970–2015, contains 552 observations of median summer flow from 37 streamgages that were operated for some periods during the 38 years from 1970 to 2015 when annual irrigation data were available, and included only the significant explanatory variables in the San Irrigation Step-Reduced Model, 1945–2015. Comparison of these two models provides a basis for assessing the robustness of parameter estimates to changes in the supporting data between the two periods. The Irrigation Step-Reduced Model, 1970–2015, uses data from 1970 through 2015, and the statistically significant variables in the San Irrigation Step-Reduced Model, 1970–2015, but augmented with irrigation data. Comparison of the San Irrigation Step-Reduced Model, 1970–2015, and the Irrigation Step-Reduced Model, 1970–2015, provides a basis for assessing if irrigation is statistically associated with water yields given the variables identified as statistically significant in the San Irrigation Step-Reduced Model, 1970–2015.

San Irrigation Full Model, 1945–2015

The San Irrigation Full Model, 1945–2015, was estimated using the data available from 1945 to 2015. Explanatory variables included the NLCD variables *WaterP*, *DevelopedP*, *BarrenP*, *ForestP*, *ShrubP*, *HerbaceousP*, *cultivatedP*, and *WetlandP*; STATSGO variables *soil_A*, *soil_B*, *soil_C*, and *soil_D*; hydrogeologic variables *K1*, *K2*, and *T2*. Time-varying characteristics included annual precipitation variables *r_ppt_jun_in* (square root of annual June precipitation, in inches), and *r_ppt_sum_in* (square root of the annual average monthly summer precipitation during July, August, and September, in inches), and the annual temperature variable *mean_TMP_degC*. In addition, a linear trend variable (*zyear*), computed as calendar year minus 1945, was included in the fixed effects. Only an intercept term for *site* was included in

the random component. The regression model included 1,478 annual median summer water yields from 40 streamgages. Maximum likelihood was used to estimate the mixed model to facilitate comparison of models with nested fixed effects but the same random effects structure.

The San Irrigation Full Model, 1945–2015, includes a constant intercept term and 19 explanatory variables as the fixed-effect component and the streamgage number for intercept deviance terms as the random component. This random component provided the flexibility for the mean response at each of the selected streamgages to differ from the overall mean response, while having the same proportional (slope) response to the fixed-effect components. Of the 19 explanatory variables included as fixed effects, 4 variables (*r_ppt_jun_in*, *r_ppt_sum_in*, *mean_TMP_degC*, and *zyear*) were highly significant (*p*-values less than 0.001), two others were significant (*p*-values less than 0.01), and three others were possibly significant (*p*-values less than 0.05), when all terms were present in the equation. Given the prevalence of statistically insignificant variables and possible correlation among parameters in this model, parameter estimates are not displayed for the full data model.

San Irrigation Step-Reduced Model, 1945–2015

Based on the San Irrigation Full Model, 1945–2015, the function “step” in package “lmerTest” was used to do a backward elimination analysis to remove nonsignificant fixed-effect variables at an alpha level of 0.01 to develop the San Irrigation Step-Reduced Model, 1945–2015. Of the 19 candidate variables, 6 were identified in the fixed-effect component as highly significant (*p*-values less than 0.001, based on the Satterthwaite approximation (Giesbrecht and Burns, 1985; Hrong-Tai Fai and Cornelius 1996) to the degrees of freedom. Selected variables in the step-reduced model include *r_ppt_jun_in*, *r_ppt_sum_in*, *mean_TMP_degC*, *soil_A*, *soil_B*, and *zyear* (table 3). The statistical significance of the random component was assessed based on a computed chi-square statistic of 349.4 based on 1 degree of freedom. The corresponding *p*-value is less than 1×10^{-7} , indicating that the null hypothesis of no statistical significance of the random component could be rejected at an alpha level of 0.01. Data used in the development of the step-reduced models are available as a USGS data release (Holtschlag, 2019).

Fixed parameter estimates for the San Irrigation Step-Reduced Model, 1945–2015 are shown in table 3. Parameter estimates are considered physically plausible with positive estimates associated with precipitation variables (*r_ppt_sum_in* and *r_ppt_jun_in*) and a negative parameter estimate for *mean_TMP_degC*. Positive parameter estimates are associated with hydrologic soil groups A and B (*soil_A* and *soil_B*), and larger parameters are associated with the soil group with lower runoff potential (U.S. Department of Agriculture, 2007). A positive trend slope associated with *zyear* is consistent with a positive trend in annual median summer water yields described previously in the report.

Table 3. Variables and parameters in the fixed-effect component of the San Irrigation Step-Reduced Model, 1945–2015, for estimating the square roots of median summer water yields in southwestern Michigan.

[*t*, a statistic formed as the ratio of the difference between estimated and expected values divided by its standard error; *t*-value, parameter estimate divided by the standard error; Prob, probability; >, greater than; *r_ppt_sum_in*, square root of the annual average monthly summer precipitation during July, August, and September, in inches; <, less than; *r_ppt_jun_in*, square root of annual June precipitation, in inches; *mean_TMP_degC*, mean summer air temperature, in degrees Celsius; *soil_A*, hydrologic soil group A; *soil_B*, hydrologic soil group B; *year*, the calendar year minus 1945]

Variable	Parameter estimate	Parameter standard error	Degree of freedom	<i>t</i> -value	Prob(> <i>t</i>)
<i>Intercept</i>	−0.1513	0.08067	197.8	−1.875	0.0623
<i>r_ppt_sum_in</i>	0.2165	0.009885	1,442	21.90	<2.0×10 ^{−16}
<i>r_ppt_jun_in</i>	0.1099	0.005699	1,443	19.28	<2.0×10 ^{−16}
<i>mean_TMP_degC</i>	−0.01684	0.002839	1,451	−5.931	3.76×10 ^{−9}
<i>soil_A</i>	0.01067	0.001346	39.70	7.928	1.05×10 ^{−9}
<i>soil_B</i>	0.006344	0.0009476	39.20	6.694	5.52×10 ^{−8}
<i>year</i>	0.001092	0.0001409	1,468	7.749	1.71×10 ^{−14}

The greatest correlation in absolute value among fixed effects was between the intercept term and *mean_TMP_degC* with a value of −0.715. The greatest correlation in absolute value among the fixed-effect variables was between *year* and *r_ppt_sum_in* with a value of −0.173. This amount of correlation is not thought to be sufficient to degrade parameter estimates.

The Akaike Information Criteria (*AIC*) of the San Irrigation Step-Reduced Model, 1945–2015, (*AICstep*=−2867.5) can be compared with *AIC* of the San Irrigation Full Model, 1945–2015, (*AICfull*=−2,738.9). Based on the equation presented by Burnham and Anderson (2004), raising Euler's number *e*, approximately 2.71828, to the power of $([AICstep - AICfull]/2)$, indicates whether the *step-reduced* or the *full* model is more likely to minimize information loss. Here, the results indicate that the San Irrigation Full Model, 1945–2015, is much less likely (about 1×10^{-28} times as likely) as the Step-Reduced Model, 1945–2015, to minimize information loss. Thus, the San Irrigation Step-Reduced Model, 1945–2015, with 6 fixed-effect variables is preferred to the San Irrigation Full Model, 1945–2015, with 19 fixed-effect variables.

The random component included intercept terms (deviances from the fixed-effect intercept) for each of the 40 streamgages selected for this analysis. The standard deviation associated random effects for *site* (intercept) was 0.08732, and the standard deviation of the residuals was 0.08715. The value of the likelihood ratio test statistic, 842, was used as the chi square statistic with one degree of freedom to evaluate the probability that the random component had no effect. Based on this computation, the probability of obtaining a value this large or larger for the statistic was less than 2.0×10^{-16} ; thus, the null hypothesis of an insignificant random component was rejected.

San Irrigation Step-Reduced Model, 1970–2015

The San Irrigation Step-Reduced Model, 1970–2015, refers to the model developed with 38 years of data from 1970 through 2015 when irrigation data was available in Michigan in addition to the other basin and climatic variables described in this report. The San Irrigation Step-Reduced Model, 1945–2015, in table 3 was re-estimated and shown in table 4 using only the 1970–2015 data to assess the sensitivity of the model to the period of record used in estimation. The San Irrigation Step-Reduced Model, 1970–2015, contains 566 observations of annual median summer water yields at 37 sites. Sites 04101800, 04102850, and 04110000 were not active when irrigation data were available. Results in tables 3 and 4 indicate that all parameters maintained their statistical significance and had comparable estimated values. Again, correlations among parameters are not thought to degrade their estimation. The *AIC* for the subset of data using the selected variates is −1,232.6.

The random component included intercept terms (deviance from the fixed-effect intercept) for each of the 37 stream sites included in this analysis. The standard deviation associated random effects for *site* (intercept) was 0.08969, and the standard deviation of the residuals was 0.07310. The value of the likelihood ratio test statistic, 382, was used as the chi square statistic to evaluate the probability that the random component had no effect. Based on this computation, the probability of obtaining a value this large or larger for the statistic is less than 2.0×10^{-16} ; thus, the null hypothesis of an insignificant random component was rejected.

Table 4. Variables and parameters in the fixed-effect component of the San Irrigation Step-Reduced Model, 1970–2015, for estimating the square roots of median summer water yields in southwestern Michigan.

[*t*, a statistic formed as the ratio of the difference between estimated and expected values divided by its standard error; *t*-value, parameter estimate divided by the standard error; Prob, probability; >, greater than; *r_ppt_sum_in*, square root of the annual average monthly summer precipitation during July, August, and September, in inches; *r_ppt_jun_in*, square root of annual June precipitation, in inches; <, less than; *mean_TMP_degC*, mean summer air temperature, in degrees Celsius; *soil_A*, hydrologic soil group A; *soil_B*, hydrologic soil group B; *zyear*, the calendar year minus 1945]

Variable	Parameter estimate	Parameter standard error	Degrees of freedom	<i>t</i> -value	Prob(> <i>t</i>)
(Intercept)	0.222001	0.115063	365.8	1.929	0.0544
<i>r_ppt_sum_in</i>	0.128172	0.017994	533.4	7.123	3.44×10^{-12}
<i>r_ppt_jun_in</i>	0.089617	0.008735	533.1	10.259	$< 2.00 \times 10^{-16}$
<i>mean_TMP_degC</i>	-0.027562	0.004247	537.7	-6.360	4.32×10^{-10}
<i>soil_A</i>	0.011067	0.001440	34.1	7.685	6.11×10^{-9}
<i>soil_B</i>	0.006366	0.001043	37.4	6.104	4.35×10^{-7}
<i>zyear</i>	0.001249	0.000264	556.3	4.724	2.93×10^{-6}

Irrigation Step-Reduced Model, 1970–2015

Estimated parameters in the Irrigation Step-Reduced Model, 1970–2015, included those in the San Irrigation Step-Reduced Model, 1970–2015, plus the irrigation variable *r_irrgMGDsqmi*. All previously estimated parameters maintained statistical significance and had comparable parameter estimated values (table 5). The sign associated with irrigation was negative, which is physically plausible, and reflects statistical significance (*p*-value=0.00531. The uncertainty of irrigation application rates, however, is not reflected in the regression, which may affect the magnitude and uncertainty of

the parameter estimate. Should more spatially precise information on basin irrigation application rates than county averages become available, re-estimation would likely produce a more reliable irrigation parameter estimate.

Given the irrigation subset data model's *AIC* of -1,238.4, the subset data model is 0.055 times as likely as the irrigation subset data model to minimize information loss, assuming basin irrigation rates were known precisely. Therefore, the Irrigation Step-Reduced Model, 1970–2015, is considered the preferred model despite the present uncertainties in the basin irrigation information.

Table 5. Variables and parameters in the fixed-effect component of the Irrigation Step-Reduced Model, 1970–2015, for estimating the square roots of median summer water yields in southwestern Michigan.

[*t*, a statistic formed as the ratio of the difference between estimated and expected values divided by its standard error; *t*-value, parameter estimate divided by the standard error; Prob, probability; >, greater than; *r_ppt_sum_in*, square root of the annual average monthly summer precipitation during July, August, and September, in inches; *r_ppt_jun_in*, square root of annual June precipitation, in inches; <, less than; *mean_TMP_degC*, mean summer air temperature, in degrees Celsius; *soil_A*, hydrologic soil group A; *soil_B*, hydrologic soil group B; *zyear*, the calendar year minus 1945; *r_irrgMGDsqmi*, square root of annual irrigation estimates, in million gallons per day per square mile]

Variable	Parameter estimate	Parameter standard error	Degrees of freedom	<i>t</i> -value	Prob(> <i>t</i>)
(Intercept)	0.240373	0.114627	350.9	2.097	0.0367
<i>r_ppt_sum_in</i>	0.122906	0.017961	538.2	6.843	2.13×10^{-11}
<i>r_ppt_jun_in</i>	0.086213	0.008756	523.9	9.846	$< 2.00 \times 10^{-16}$
<i>mean_TMP_degC</i>	-0.027883	0.004228	539.5	-6.595	1.01×10^{-10}
<i>soil_A</i>	0.012180	0.001494	38.2	8.151	6.88×10^{-10}
<i>soil_B</i>	0.006684	0.001049	38.1	6.368	1.76×10^{-7}
<i>zyear</i>	0.001649	0.000299	562.9	5.516	5.30×10^{-8}
<i>r_irrgMGDsqmi</i>	-0.349441	0.124805	490.0	-2.800	0.00531

The random site-intercept deviances with 95-percent confidence intervals under the irrigation subset data model are shown on figure 6. The widths of these confidence intervals monotonically decrease with increasing years of record used in the analysis for each site. The standard deviation of the random site-intercept deviances themselves was 0.08981, and the standard deviation of the residuals was 0.07256.

The value of the likelihood ratio test statistic, 369, was used as the chi square statistic to evaluate the probability that the random component had no effect. Based on this computation, the probability of obtaining a value this large or larger for the statistic is less than 2.0×10^{-16} ; thus, the null hypothesis of an insignificant random component was rejected.

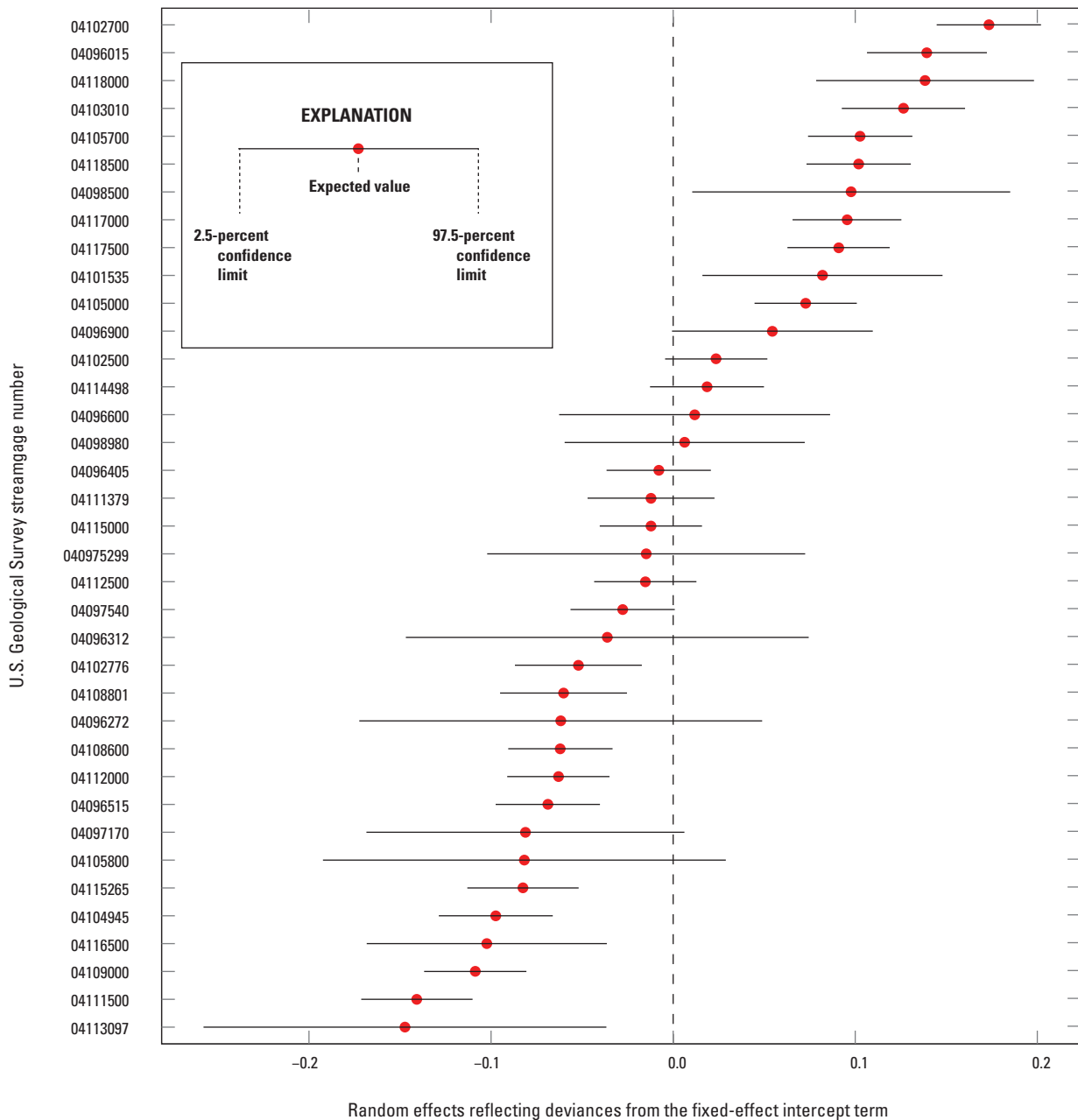


Figure 6. Random effects reflecting deviances from the fixed-effect intercept term with 95-percent confidence intervals.

The relation between measured and fixed plus random components predicted water yields for the irrigation subset data model is shown in figure 7. For mixed models, R^2 computed on the basis of both fixed and random effects is referred to as conditional R^2 , whereas R^2 computed on the basis of fixed effects only is referred to as the marginal R^2 . The conditional and marginal R^2 values for the irrigation subset data model are

0.8664 and 0.6617, respectively, based on the function “r.squaredGLMM” in the R package “MuMIn” (Barton, 2017). Normality of residuals was not rejected based on a D statistic of 0.033299 and a p -value of 0.134 at the alpha level of 0.05, based on the Lilliefors test applied using function “lillie.test” in the R package “nortest” (Gross and Ligges, 2015).

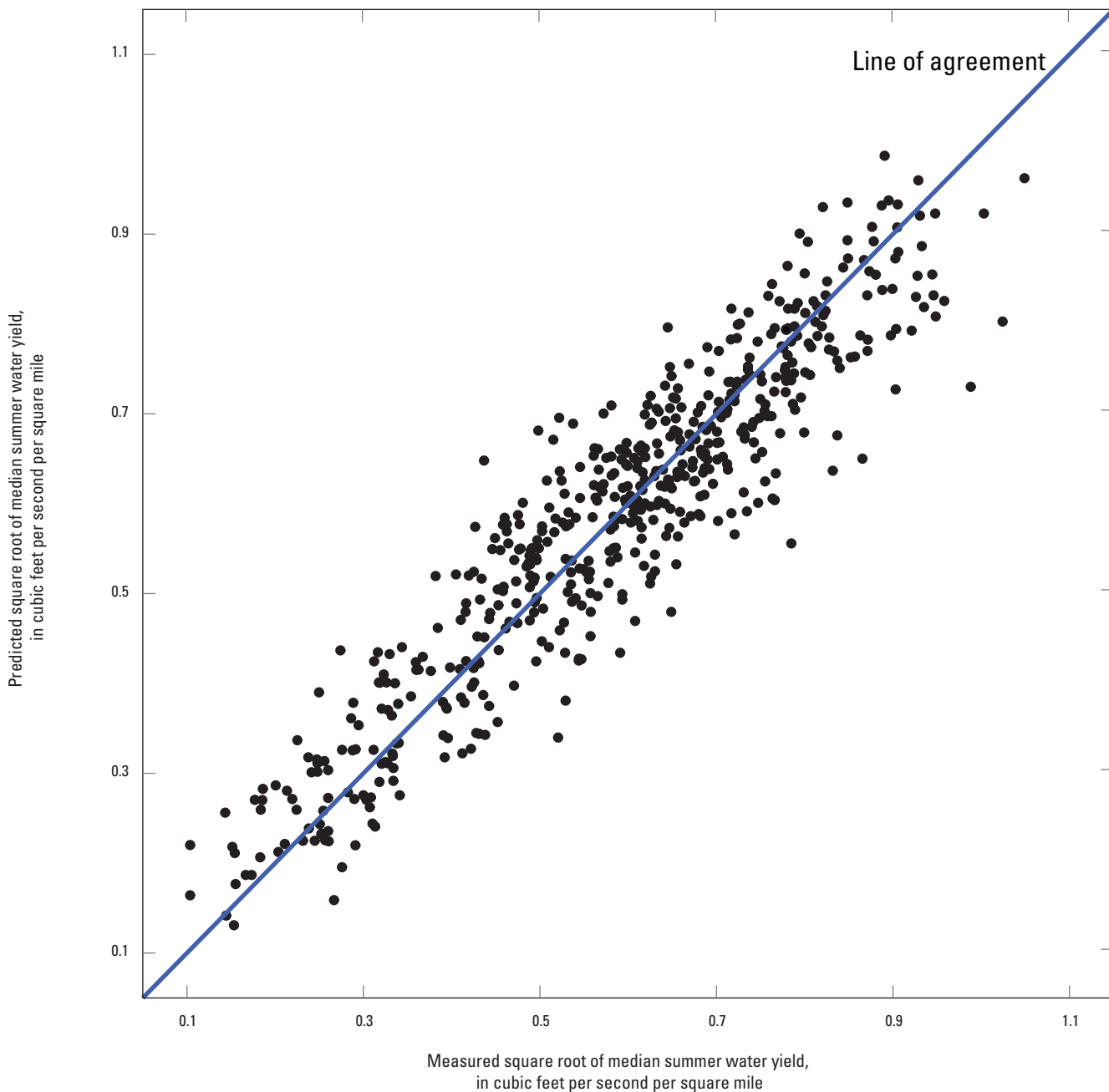


Figure 7. Relation between measured and predicted median summer water yields based on the Irrigation Step-Reduced Model, 1970–2015, for selected streamgages in southwestern Michigan.

Estimation of Median Summer Water Yields at Streamgages in Southwestern Michigan

The mixed models developed in this report may be used to estimate the magnitudes and uncertainties of annual median water yields at the selected USGS streamgages used in this report and at non-selected sites in southwestern Michigan that have basin, climatic, and irrigation characteristics like the basins used in model development. In general, this would restrict application to small, rural basins having unregulated flows. Model estimates would be inverse transformed (squared) and multiplied by the corresponding drainage area of the target basin upstream of the streamgage of interest to estimate annual median flows.

The irrigation subset data model will be used to illustrate the estimation process, although the process is applicable to other mixed models developed in the report. Model application would require estimates of the explanatory variables, with appropriate transformations. For estimates where random effects are not known (all sites not used in model development, which are referred to as nonselected sites), an estimate of the square root of water yield can be computed using parameter estimates in table 5. If a prediction interval is needed, the “predictInterval” function in the “merTools” R package (Knowles and Frederick, 2016) can be used with the model object as described in the following example.

Within the R environment, the script shown in figure 8 can be used to compute estimates and prediction intervals using the fixed and random components for selected sites at streamgages used in this report and the fixed components only for other nonselected sites. Note that the line numbers on the left-hand side are for reference only, and are not part of the R code.

Starting in line 3, any needed uninstalled R packages are installed, and packages are loaded with the library functions to enable the computation. In line 12, the mixed model object is loaded into the R environment with the “load(file=‘lmer_FXD_ppt_TMP_Trnd_SoilA_B_Irrg_RND_site_DATA_sub.RData’)” function; the mixed model object is available electronically for download. Once the model object is loaded, a data frame for study site USGS streamgage 04102700 is created using corresponding data in this report, as shown in lines 15–18. The selection of streamgage 04102700 was arbitrary among the 40 streamgages used in the analysis to illustrate the computational approach. The random number seed is set in line 19 to facilitate reproducibility of results because the Monte Carlo simulations used to develop prediction intervals involve random numbers. In lines 21–24, the “predictInterval” function is used to compute the expected value (*fit*), and upper (*upr*) and lower (*lwr*) confidence limits, which are stored in the data frame *pred_study_site*. The *level* parameter specifies 68.29 percent of the total area is under the standard normal distribution, which corresponds to plus or minus 1 standard deviation.

In line 29, a second, arbitrary site was specified as ‘non-select_04102700’, which had basin, climate, and irrigation factors equal to those at the site where streamgage 04102700 is located, but it has different site identifier. Therefore, only the “site” identifier is revised, and then the data frame is used on lines 31–34 to compute a new expected value (*fit*) and confidence limits for the nonselected site. When the “predictInterval” function is executed for a nonselected site, a warning message is displayed: “The following levels of site from *newdata* [site identifier], – are not in the model data. Currently, predictions for these values are based only on the fixed coefficients and the observation-level error.” The warning message implies that there is no site-specific deviance for nonselected sites, so the site deviance is set to zero. Also, the current version of the “predictInterval” function underestimates the variance of the prediction because the increase in variance associated with the unknown random intercept at a nonselected site is not estimated.

Until this issue in the “predictInterval” function is addressed by the package developer, the estimated prediction interval can be improved by including the uncertainty of the random intercept at a nonselected site. Assuming a nonselected site is selected at random from the same population of selected sites sampled in model development, the standard deviation of the random intercept (line 38) can be computed as 0.08981. The adjusted standard deviation is computed as the square root of the variance of fixed coefficients plus the variance of the intercept deviances (line 40). The estimated value of *fit* for the new site is used with the adjusted standard deviation to compute the expected distribution of the prediction.

The distribution of estimates based on the fixed and random components of the irrigation subset data model for selected site at USGS streamgage 04102700, and a site having the same characteristics but a different identifier, are shown in figure 9. The distributions for the two site estimates differ in their expected values and variances. The expected values differ by the random intercept estimated for selected site 04102700 of 0.18 cubic foot per second per square mile. Thus, the expected value for selected site 04102700 is about 0.71 cubic foot per second per square mile, whereas an expected value for nonselected site 04102700 is about 0.53 cubic foot per second per square mile. The mean random intercept, constrained to zero, is applied to all nonselected sites. The variances differ because of the uncertainty of the random intercept for nonselected site 04102700 is unknown but estimated as the variance of all random intercepts estimated for the selected sites.

```

1  ### Load the Irrigation Subset Data Mixed Model Object and predict density
2  ```{r predict_sqrt_yields, echo = TRUE}
3  # Load packages used in analysis (May need to install packages prior to loading.)
4  need_pak <- c('merTools', 'lmerTest', 'Matrix', 'lme4', 'tidyverse', 'ggthemes')
5  have_pak <- rownames(installed.packages())
6  if (length(setdiff(need_pak, have_pak)) > 0 ) {
7    install.packages( setdiff(need_pak, have_pak) )
8  }
9  library(merTools); library(lmerTest); library(Matrix); library(lme4); library(ggplot2);
10 library(ggthemes)
11 # Irrigation subset data mixed model object
12 load(file = 'lmer_FXD_ppt_TMP_Trnd_SoilA_B_Irrg_RND_site_DATA_sub.RData')
13 # Define a data frame containing explanatory variables
14 # Note: zyear = 70 corresponds to year 2015
15 df1 <- data.frame(site = '04102700', r_ppt_jun_in_wt = 2.367911,
16                   r_ppt_sum_in_wt = 1.806103, wt_mean_TMP_degC = 20.11033,
17                   soil_A = 7.019279, soil_B = 42.75795, zyear = 70,
18                   r_irrgMGDsqmi = 0.16827373)
19 set.seed(123456) # set random number seed for reproducibility of results
20 # predict square root of water yields for select site at USGS streamgage 04102700
21 pred_select_site <- predictInterval(lmer_FXD_ppt_TMP_Trnd_SoilA_B_Irrg_RND_site_DATA_sub,
22                                   newdata = df1, which = "full",
23                                   level = 0.6826, n.sims = 5000, stat = "mean",
24                                   type = "linear.prediction")
25
26 # Difference between between fit and lwr is 1 standard deviation because level = 0.6826
27 pred_select_site_sd <- pred_select_site$fit - pred_select_site$lwr
28 # Change to a nonselect site name with explanatory variables equal to 04102700
29 df1$site = 'nonselect_04102700'
30 # predict square root of water yields for site not in select
31 pred_nonselect_site <- predictInterval(lmer_FXD_ppt_TMP_Trnd_SoilA_B_Irrg_RND_site_DATA_sub,
32                                       newdata = df1, which = "full",
33                                       level = 0.6826, n.sims = 5000, stat = "mean",
34                                       type = "linear.prediction")
35 # Difference between between fit and lwr is 1 standard deviation because level = 0.6826
36 pred_nonselect_site_sd <- pred_nonselect_site$fit - pred_nonselect_site$lwr
37 # Compute standard deviation of random effects (intercept deviations)
38 re_sd <- sd(ranef(lmer_FXD_ppt_TMP_Trnd_SoilA_B_Irrg_RND_site_DATA_sub)$site$(Intercept))
39 # Add variances to fixed and random effects
40 pred_nonselect_site_sd <- sqrt(pred_nonselect_site_sd^2 + re_sd^2)
41 # Plot figure 9
42 ggplot(data = data.frame(x = c(0.1, 1.1)), aes(x)) +
43   stat_function(fun = dnorm, n = 201, args = list(mean = pred_select_site$fit,
44         sd = pred_select_site_sd), color = 'red') +
45   stat_function(fun = dnorm, n = 201, args = list(mean = pred_nonselect_site$fit,
46         sd = pred_nonselect_site_sd), color = 'blue') +
47   xlab('Predicted Square root of water yield, in cubic feet per second per square mile') +
48   ylab('Probability density') +
49   geom_vline( xintercept = pred_select_site$fit, color = 'red', linetype = 'dashed') +
50   geom_vline( xintercept = pred_nonselect_site$fit, color = 'blue', linetype = 'dashed') +
51   annotate('text', x = pred_select_site$fit + 0.066, y = 5.7, label = 'select_04102700') +
52   annotate('text', x = pred_nonselect_site$fit + 0.075, y = 5.5, label =
53     'nonselect_04102700') +
54   scale_y_continuous(limits = c(0,6), breaks = seq(0,6)) +
55   scale_x_continuous(limits = c(0.1, 1.1), breaks = seq(0.1, 1.1, 0.1)) +
56   theme_few() +
57   theme(axis.text.x = element_text(margin=unit(c(0.5, 0.5,0.5,0.5), "cm"), size = 12),
58         axis.text.y = element_text(margin=unit(c(0.5, 0.5,0.5,0.5), "cm"), size = 12),
59         axis.ticks.length=unit(-0.20, "cm"))
60 ...

```

Figure 8. R script to compute the magnitude and prediction intervals for the square root of median summer water yields in southwestern Michigan streams using mixed model object and specified data.

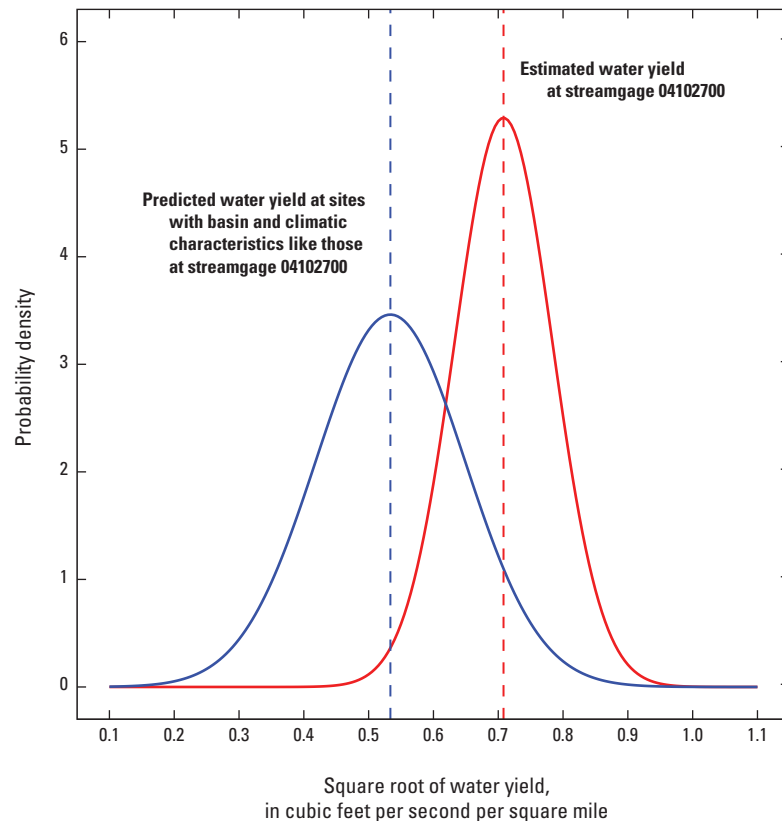


Figure 9. Distribution of estimated median summer water yields for U.S. Geological Survey streamgage 04102700 and other stream sites with the same basin and climatic characteristics as those at 04102700.

Summary

Forty U.S. Geological Survey streamgages in southwestern Michigan were selected to investigate median flow conditions during the summer months of July, August, and September. Selected streamgages monitored flow from mainly rural basins less than 1,000 square miles, where effects of regulation or storage were thought to have limited effects on median summer flows. Analysis included annual flow data from 1945 through 2015, a period when seven or more streamgages were operational in any given year. To facilitate comparison among basins of different sizes, annual summer water yields, computed by dividing annual median summer flows by corresponding basin drainage areas, were considered the response variable. A square root transformation was applied to water yields to normalize their distribution before model development.

The distributions of median summer flows as water yields were compared with flows for 27 streamgages in southwestern Michigan that had been used in an earlier study of median

summer flows. Index flows are defined as the median flow during the summer month of lowest flow during the period of record available at the streamgage. The Michigan Department of Environmental Quality uses index flow information to help prevent negative effects to characteristic fish populations from large water withdrawals, which are commonly associated with irrigation in rural areas.

Index water yields were generally between the lower quartile (25th percentile) and the median annual median summer flow, never exceeding the upper quartile (75th percentile). The index water yield corresponds, on average, to the 0.37 quantile of annual median water yield. Use of an annual statistic like an annual median summer water yield, rather than a period-of-record statistic like index water yield, provides a basis for understanding how annual precipitation and irrigation may change over time and affect median summer water yields.

Mixed models were developed to identify time-invariant and time-varying explanatory variables that are statistically associated with the annual series of median summer water yields. A mixed model is an extension of an ordinary least

square regression model that includes fixed and random components. The fixed component collectively describes the expected (or fixed effect) relation between individual selected explanatory variables and water yields at all applicable sites. The random component describes site-specific variations in specified fixed effects. Possible explanatory variables included sets of time-invariant basin characteristics and sets of time-varying climatic and irrigation variables. Basin characteristics included major land cover characteristics, primary hydrologic soil groups, and measures of hydraulic conductivity and transmissivity of underlying geologic materials. Time-varying characteristics included monthly precipitation and temperature data and annual irrigation data aggregated to the county level. A linear trend term also was included as a possible explanatory variable.

Four mixed models were developed to identify explanatory variables associated with annual series of median summer water yields as fixed effects. All mixed models included a random intercept term. The full data model was estimated using data from 1945 to 2015, which included 1,478 values of summer water yields, 19 explanatory variables as the fixed component, and a site label, corresponding to the streamgage number, as a random component. The step-reduced model used backward elimination process to reduce the number of fixed effects from 19 to 6 variables including: *r_ppt_jun_in* (square root of annual June precipitation, in inches), *r_ppt_sum_in* (square root of annual average summer precipitation during July, August, and September, in inches), *mean_TMP_degC* (mean summer air temperature, in degrees Celsius), *soil_A*, *soil_B* (percentage of basin area with hydrologic soil groups A and B, respectively), and *zyear* (calendar year minus 1945). The step-reduced model was re-estimated by the subset data model containing 552 values of median summer water yield for 38 years within the period 1970–2015 when irrigation information was available. All parameters of the step-reduced model estimated with the subset data model maintained their statistical significance and their algebraic signs and were of comparable magnitudes with those in the step-reduced model. Lastly, an irrigation subset data model was estimated using the 552 values of median summer water yields with annual basin irrigation (square root of annual irrigation estimates, in million gallons per day per square mile [*r_irrgMGDsqmi*]) data as a fixed effect. The parameter associated with irrigation was negative and statistically improved the mixed model. The process for estimating the magnitude and uncertainty of median summer water yields by use of the irrigation subset data model at a selected site used to develop the model, and at a nonselected site, was demonstrated. The irrigation subset data model might be improved by re-estimating model parameters using more spatially precise information about irrigation withdrawals.

References Cited

- Anderson, J.R., Hardy, E.E., Roach, J.T., and Witmer, R.E., 1976, A land use and land cover classification system for use with remote sensor data: U.S. Geological Survey Professional Paper 964, 28 p.
- Barton, K., 2017, MuMIn—Multi-Model Inference (ver. 1.40.0): R package, accessed December 8, 2017, at <https://CRAN.R-project.org/package=MuMIn>.
- Bates, D., Maechler, M., Bolker, B.M., and Walker, S.C., 2015, Fitting linear mixed-effects models using lme4: Journal of Statistical Software, v. 67, no. 1, 48 p., accessed April 2018 at <https://doi.org/10.18637/jss.v067.i01>.
- Bayless, E.R., Arihood, L.D., Reeves, H.W., Sperl, B.J.S., Qi, S.L., Stipe, V.E., and Bunch, A.R., 2017, Maps and grids of hydrogeologic information created from standardized water-well drillers' records of the glaciated United States: U.S. Geological Survey Scientific Investigations Report 2015–5105, 34 p., accessed April 2018 at <https://doi.org/10.3133/sir20155105>.
- Bedell, D.J., and Van Til, R.L., 1979, Irrigation in Michigan, 1977: Lansing, Mich., Michigan Department of Natural Resources, Water Management Division, 37 p.
- Burnham, K.P., and Anderson, D.R., 2004, Multimodel inference—Understanding AIC and BIC in model selection: Sociological Methods & Research, v. 33, no. 2, p. 261–304, accessed April 2018 at <https://doi.org/10.1177/0049124104268644>.
- Draper, N.R., and Smith, H., 1998, Applied regression analysis (3d ed.): New York, John Wiley, 736 p.
- Giesbrecht, F.G., and Burns, J.C., 1985, Two-stage analysis based on a mixed model—Large-sample asymptotic theory and small-sample simulation results: Biometrics, v. 41, no. 2, p. 477–486, accessed April 2018 at <https://doi.org/10.2307/2530872>.
- Gross, J., and Ligges, U., 2015, nortest—Tests for Normality (ver. 1.0-4): R package, accessed December 8, 2017, at <https://CRAN.R-project.org/package=nortest>.
- Hamilton, D.A., Sorrell, R.C., and Holtschlag, D.J., 2008, A regression model for computing index flows describing the median flow for the summer month of lowest flow in Michigan: U.S. Geological Survey Scientific Investigations Report 2008–5096, 43 p. [Also available at <https://pubs.usgs.gov/sir/2008/5096/>.]

- Hodgkins, G.A., Dudley, R.W., and Aichele, S.S., 2007, Historical changes in precipitation and streamflow in the U.S. Great Lakes Basin, 1915–2004: U.S. Geological Survey Scientific Investigations Report 2007–5118, 31 p. [Also available at <https://pubs.usgs.gov/sir/2007/5118/>.]
- Holtschlag, D.J., 2019, Data on factors affecting spatial and temporal variations of annual summer median water yields in southwestern Michigan, 1945–2015: U.S. Geological Survey data release, <https://doi.org/10.5066/P9WNIF7M>.
- Homer, C.G., Dewitz, J.A., Yang, L., Jin, S., Danielson, P., Xian, G., Coulston, J., Herold, N.D., Wickham, J.D., and Megown, K., 2015, Completion of the 2011 National Land Cover Database for the conterminous United States—Representing a decade of land cover change information: Photogrammetric Engineering and Remote Sensing, v. 81, no. 5, p. 345–354.
- Hrong-Tai Fai, A., and Cornelius, P.L., 1996, Approximate F-tests of multiple degree of freedom hypotheses in generalized least squares analyses of unbalanced split-plot experiments: Journal of Statistical Computation and Simulation, v. 54, no. 4, p. 363–378., accessed April 2018 at <https://doi.org/10.1080/00949659608811740>.
- Hutson, S.S., comp., 2007, Guidelines for preparation of State water-use estimates for 2005: U.S. Geological Survey Techniques and Methods, book 4, chap. E1, 28 p. [Also available at <https://pubs.usgs.gov/tm/2007/tm4e1/>.]
- Iman, R.L., and Conover, W.J., 1983, A modern approach to statistics, New York, John Wiley & Sons, 524 p.
- Knowles, J.E., and Frederick, C., 2016, merTools—Tools for analyzing mixed effect regression models (ver. 0.3.0): R package, accessed December 8, 2017, at <https://cran.r-project.org/package=merTools>.
- Kuznetsova, A., Brockhoff, P.B., and Christensen, R.H.B., 2016, lmerTest—Tests in linear mixed effects models (ver. 2.0-33): R package, accessed December 5, 2017, at https://cran.r-project.org/src/contrib/Archive/lmerTest/lmerTest_2.0-33.tar.gz.
- McKay, L., Bondelid, T., Dewald, T., Johnston, J., Moore, R., and Rea, A., 2017, NHDPlus version 2—User guide: U.S. Environmental Protection Agency, 179 p., accessed November 30, 2017, at ftp://ftp.horizon-systems.com/NHD-plus/NHDPlusV21/Documentation/NHDPlusV2_User_Guide.pdf.
- Michigan Department of Environmental Quality, 2018, Michigan water use data: Michigan Department of Environmental Quality, digital data, accessed March 2, 2018, at http://www.michigan.gov/deq/0,4561,7-135-3313_3684_45331-370128-,00.html.
- Michigan Department of Environmental Quality, Michigan Department of Agriculture, and Michigan State University, 2018, Historical Michigan water use reports, data and graphics: Michigan Department of Environmental Quality, digital data, accessed March 2, 2018, at http://www.michigan.gov/deq/0,4561,7-135-3313_3684_45331-381590-,00.html. [Downloaded files: [yyyy] Agricultural Irrigation.pdf, where [yyyy] indicates the four-digit calendar year.]
- Morenz, M.L., Van Til, R.L., and Luukkonen, C.L., 2005, Water use for irrigation in Michigan, 2001: U.S. Geological Survey Fact Sheet 2005–3079, 4 p., accessed December 5, 2017, at <https://pubs.water.usgs.gov/FS2005-3079/>.
- R Core Team, 2017, R—A language and environment for statistical computing (ver. 3.4.3): Vienna, Austria, R Foundation for Statistical Computing, software, accessed December 2017 at <https://www.R-project.org/>.
- Seaber, P.R., Kapinos, F.P., and Knapp, G.L., 1994, Hydrologic unit maps: U.S. Geological Survey Water-Supply Paper 2294, 63 p. [Also available at <https://pubs.usgs.gov/wsp/wsp2294/>.]
- State of Michigan, 2006, Enrolled Senate bill no. 850, act no. 33 of Public acts of 2006: Michigan Legislature, 93d Legislature, accessed November 28, 2017, at <http://www.legislature.mi.gov/documents/2005-2006/publicact/pdf/2006-PA-0033.pdf>.
- U.S. Department of Agriculture, 1994, State Soil Geographic (STATSGO) database: Lincoln, Nebr., Natural Resources Conservation Service, National Soil Survey Center, Miscellaneous Publication No. 1492, 113 p.
- U.S. Department of Agriculture, 2007, Hydrologic soil groups, chap. 7 of National Engineering Handbook: Natural Resources Conservation Service, pt. 630, 5 p., accessed December 5, 2017, at <https://directives.sc.egov.usda.gov/OpenNonWebContent.aspx?content=17757.wba>.
- U.S. Environmental Protection Agency, and U.S. Geological Survey, 2012, NLCD 2011 land use extension: Horizon Systems Corporation website, NHDPlusV2 digital data, accessed December 11, 2017, at <http://www.horizon-systems.com/NHDPlus/V2NLCD2011.php#download>. [Downloaded file: NHDPlusV21_GL_04_VPUAttributeExtensionNLCD_01.7z.]
- U.S. Geological Survey, 2017, Water use data for Michigan, in Water use data for the Nation: U.S. Geological Survey, National Water Information System database, accessed February 26, 2017, at <https://doi.org/10.5066/F7P55KJN>. [State information directly accessible at https://waterdata.usgs.gov/mi/nwis/water_use/.]

- Wieczorek, M.E., Jackson, S.E., and Schwarz, G.E., 2016, Attributes for NHDPlus Version 2.1 reach catchments and modified routed upstream watersheds for the conterminous United States—STATSGO soil characteristics, hydrologic groups: U.S. Geological Survey digital data, accessed November 30, 2017, at <https://www.sciencebase.gov/catalog/item/5728d93be4b0b13d3918a99f>. [Downloaded data: STATSGO_HYDGRP_CAT_CONUS.zip.]
- Wieczorek, M.E., Wolock, D.M., Jackson, S.E., and Schwarz, G.E., 2017a, Attributes for NHDPlus Version 2.1 reach catchments and modified-routing upstream watersheds for the conterminous United States—Monthly average precipitation (millimeters) from 1945–2015: U.S. Geological Survey digital data, accessed December 4, 2017, at <https://www.sciencebase.gov/catalog/item/5730cf1ce4b0dae0d5db1f17>. [Downloaded files: PPT[yyyy]_CAT_CONUS.zip, where [yyyy] indicates the four-digit calendar year, within the “Catchment Values” Child Item.]
- Wieczorek, M.E., Wolock, D.M., Jackson, S.E., and Schwarz, G.E., 2017b, Attributes for NHDPlus Version 2.1 reach catchments and modified-routing upstream watersheds for the conterminous United States—Monthly average temperature (Celsius) from 1945–2015: U.S. Geological Survey digital data, accessed October 12, 2017, at <https://www.sciencebase.gov/catalog/item/5730cf70e4b0dae0d5db1f1c>. [Downloaded files: TAV[yyyy]_CAT_CONUS.zip, where [yyyy] indicates the four-digit calendar year, within the “Catchment Values” Child Item.]

For additional information contact:

[Director, Upper Midwest Water Science Center](#)

U.S. Geological Survey

6520 Mercantile Way

Suite 5

Lansing, MI 48911

Publishing support provided by the

Rolla and Madison Publishing Service Centers

