Prepared in cooperation with the Maine Department of Transportation

Peak Flow Regression Equations for Small, Ungaged Streams in Maine: Comparing Map-Based to Field-Based Variables

Scientific Investigations Report 2015–5049

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

Peak Flow Regression Equations for Small, Ungaged Streams in Maine: Comparing Map-Based to Field-Based Variables

By Pamela J. Lombard and Glenn A. Hodgkins

Scientific Investigations Report 2015–5049

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

U.S. Department of the Interior
SALLY JEWELL, Secretary

U.S. Geological Survey
Suzette M. Kimball, Acting Director

U.S. Geological Survey, Reston, Virginia: 2015

Acknowledgments

The authors wish to thank Alexander Mann at the Maine Department of Transportation for his work in delineating drainage basins by using orthoimagery.

Thanks are also given to Terrence Talbot at the U.S. Geological Survey for his work collecting field data and building HEC-RAS models.

Contents

Figures

Tables

Conversion Factors

Inch/Pound to SI

Multiply	By	To obtain
Length		
inch (in.)	2.54	centimeter (cm)
foot (ft)	0.3048	meter (m)
mile (mi)	1.609	kilometer (km)
Area		
square foot (ft^2)	929.0	square centimeter (cm^2)
square foot (ft^2)	0.09290	square meter (m^2)
square inch (in^2)	6.452	square centimeter (cm^2)
square mile (mi^2)	2.590	square kilometer (km^2)
Flow rate		
cubic foot per second (ft^3/s)	0.02832	cubic meter per second (m^3/s)

Peak Flow Regression Equations for Small, Ungaged Streams in Maine: Comparing Map-Based to Field-Based Variables

By Pamela J. Lombard and Glenn A. Hodgkins

Abstract

Regression equations to estimate peak streamflows with 1- to 500-year recurrence intervals (annual exceedance probabilities from 99 to 0.2 percent, respectively) were developed for small, ungaged streams in Maine. Equations presented here are the best available equations for estimating peak flows at ungaged basins in Maine with drainage areas from 0.3 to 12 square miles (mi^2). Previously developed equations continue to be the best available equations for estimating peak flows for basin areas greater than 12 mi^2. New equations presented here are based on streamflow records at 40 U.S. Geological Survey streamgages with a minimum of 10 years of recorded peak flows between 1963 and 2012. Ordinary least-squares regression techniques were used to determine the best explanatory variables for the regression equations. Traditional map-based explanatory variables were compared to variables requiring field measurements. Two field-based variables—culvert rust lines and bankfull channel widths—either were not commonly found or did not explain enough of the variability in the peak flows to warrant inclusion in the equations. The best explanatory variables were drainage area and percent basin wetlands; values for these variables were determined with a geographic information system. Generalized least-squares regression was used with these two variables to determine the equation coefficients and estimates of accuracy for the final equations.

Introduction

The Maine Department of Transportation (MaineDOT) and other State and town agencies tasked with the construction and maintenance of drainage infrastructure in or near rivers and streams need methods for estimating streamflows at ungaged locations on small streams throughout Maine. Equations for estimating peak-flow statistics at streams with 1-, 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals (equivalent to annual exceedance probabilities

[AEPs] of 99-, 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent, respectively) can be useful for culvert design. A flood with a 1-percent AEP has a 1-in-100 chance of being equaled or exceeded in any given year and has an average recurrence interval of 100 years and is referred to as the 100-year flood. Statistical methods currently are available to estimate peak flows for ungaged streams at basins in Maine with drainage areas from 1 to 1,700 square miles (mi^2) (Hodgkins, 1999) and monthly mean and median flows at basins with drainage areas from 10 to 1,500 mi^2 (Dudley, 2004). Methods are not available, however, for estimating peak flows for very small basins (drainage areas less than 1 mi^2) or for estimating mean and median monthly flows for small basins (drainage areas between 1 and 10 mi^2). The exception is for June and August median flows in some parts of the State as documented in Lombard and others (2003) and Lombard (2004, 2010). Although current peak-flow regression equations can be used for basins with drainage areas as small as 1 mi^2, the resulting estimates may not be robust for basins this small because the equations were developed using few sites with drainage areas less than 3 mi^2. In 2011, the U.S. Geological Survey began a cooperative investigation with the MaineDOT to test regression-equation explanatory variables for estimating peak flows on small (less than 12 mi^2) basins in Maine.

Previous Studies

In 2007, the accuracy of various methods used for estimating peak flows for small and very small drainage basins in Maine was explored (Hodgkins and others, 2007). For very small basins, the simple hydrologic models TR-20 and the Rational Method were found to have very large errors (root mean square errors of –98.0 to 5,010 percent and –69.7 to 230 percent, respectively). It was not possible to test regression equations for estimating peak flows on very small basins due to the lack of very small basins with sufficient data. Regression equations were found to be more accurate than TR-20, the Rational Method, and the Probabilistic Rational Method for estimating peak flows at small basins (Hodgkins and others, 2007).

The possibility for testing regression equations for computing peak flows at very small, ungaged basins now exists because of data collected in the last decade by the U.S. Geological Survey (USGS) in cooperation with MaineDOT and others. In addition, easy-to-obtain field indicators of peak streamflows at culverts, such as culvert rust lines and bankfull channel geometry, may allow for improved accuracy of regression equations for estimating peak flows at small basins in Maine. Many culvert rust lines have been observed by MaineDOT that seem to correspond with high flows through the culverts. Bankfull channel geometry often corresponds to defined high flows (Leopold and Maddock, 1953; Dunne and Leopold, 1978) and may be a good explanatory variable for peak flows.

Purpose and Scope

The purpose of this report is to present the results of a study conducted to develop new regression equations for estimating peak flows at small (less than 12 mi^2), ungaged streams in Maine and to determine if field indicators can improve the accuracy of these equations. The objectives of the study described in this report include (1) computing peak flows for selected recurrence intervals at all small, gaged basins in Maine, including using peak-flow data that were collected during the last 15 years; (2) determining whether two field indicators of peak flows—culvert rust lines and bankfull channel geometry—are generally available and improve the accuracy of peak-flow regression equations compared with equations developed solely using basin characteristics such as drainage area; and (3) comparing the best new equations for small basins with previously published statewide peak-flow equations.

Data Collection

Data collection for this study involved determining which small drainage basins in Maine have both streamflow gaging data and usable field indicators, surveying and measuring channel and culvert geometry at usable sites, and using a geographic information system (GIS) to delineate basins and calculate basin characteristics for each site. Sufficient field data were collected to build hydraulic models for sites with established culvert rust lines. The basin characteristics were analyzed to aid in the development of regression equations for computing flood frequency at ungaged sites.

Site Identification and Surveying of Field-Based Variables

Forty USGS streamgaging stations in Maine with drainage areas less than 12 mi^2 and a minimum of 10 years of annual peak-flow record (table 1; fig. 1) were used in the study. Note that 10 years of record is generally considered

the minimum amount of data needed for estimating peak flows on small streams. Flood-frequency statistics would be expected to be more accurate at stations with longer periods of record. Eight of the 40 stations have continuous daily flow data, and the remaining 32 stations have annual peak-flow data that were measured with crest-stage gages, a cost-effective method of measuring peak flows that exceed a set threshold. Field reconnaissance at each of the 40 sites determined that 23 sites had measurable geomorphic indicators of bankfull width (fig. 2), and 7 sites had well-defined culvert rust lines (fig. 3). The majority of these sites were located near culverts, although only seven had well-defined rust lines.

Bankfull width was determined at locations with a clear break between the channel and the flood plain, well-defined changes in vegetation, and (or) the top of the inside bend of a point bar (alluvium that has accumulated on the inside bend of a meandering stream). The first two indicators were the most common. Five measurements of bankfull width were recorded at each of the sites where bankfull width could be adequately defined. The five widths were taken at locations along the channel that were a minimum distance of 10 culvert diameters upstream or downstream from the culvert and a maximum distance of 1,000 feet (ft) upstream or downstream from the culvert. The width measurements along the channel were made at a distance of at least one culvert diameter from each other. At the locations where bankfull widths were measured, there were no issues with unstable banks, exposed bedrock, nearby boulders, artificial encroachments, channel obstructions, braided channels, or backwater from downstream constrictions (which could include the nearby culvert). Bankfull widths were not measured near falls or sharp bends. A median bankfull width was computed from the five measured widths at each site; median bankfull widths near the 23 culverts ranged from 6.8 to 32.3 ft (table 1).

For sites with well-defined culvert rust lines, culvert-geometry and cross-section data were obtained from field surveys conducted by personnel from the USGS New England Water Science Center during the summer of 2012. Surveys used differential leveling techniques to tie elevations to a local datum. Characteristics of each culvert were determined, and structural dimensions were measured.

Map-Based Determination of Basin Characteristics

Basin characteristics were determined using a GIS. Basin delineation errors can result in large overall errors in basin characteristics, especially for sites with small basins. Thus, it was especially important to have accurate measurements of drainage areas for this study. Several techniques were used to obtain drainage areas on the small sites. In 12 Maine basins for which light detection and ranging (lidar) data were available in 2014, a digital elevation model (DEM) derived from the lidar data was used to delineate the basins by using ArcHydro within ArcGIS. ArcHydro is a set of data models and tools that

Table 1. Basin characteristics at 40 USGS streamgaging stations on small streams in Maine.

[USGS, U.S. Geological Survey; mi², square miles; ME, Maine; NA, not available. Crest stage gages measure peak flow only]

USGS streamgage number	Streamgage name	Gage type	Years of record	Drainage area (mi²)	Basin wetlands (percent)	Average basin slope (percent)	Median bankfull width (feet)
01012895	Clark Brook near Wallagrass Plantation, ME (Eagle Lake)	Crest stage	13	1.02	10.6	55.4	NA
01012970	Unnamed Brook near Wallagrass Plantation, ME (Soldier Pond)	Crest stage	13	0.42	1.60	38.3	8.6
01014700[1]	Factory Brook near Madawaska, ME	Crest stage	11	5.97	2.01	38.2	25.0
01015700	Houlton Brook near Oxbow, ME	Crest stage	11	6.15	20.9	11.3	17.0
01017300	Nichols Brook near Caribou, ME	Crest stage	11	3.87	7.41	14.9	NA
01017550	Williams Brook at Phair, ME	Continuous	14	3.62	22.2	11.5	NA
01017900	Marley Brook near Ludlow, ME	Continuous	18	1.57	0.88	19.0	NA
01021300	Wiggins Brook near West Lubec, ME	Crest stage	10	4.98	11.6	19.3	NA
01021470	Libby Brook near Northfield, ME	Continuous	13	6.73	13.4	10.2	NA
01021600	Middle River near Machias, ME	Crest stage	10	8.05	14.0	17.7	19.2
01021890	Unnamed Brook near Crawford, ME	Crest stage	13	0.70	5.57	19.6	NA
01022700	Forbes Pond Brook near Prospect Harbor, ME	Crest stage	10	8.53	20.1	19.4	NA
01024200	Garland Brook near Mariaville, ME	Continuous	18	9.75	6.68	29.9	32.3
01026800	Frost Pond Brook near Sedgwick, ME	Crest stage	10	6.26	17.9	22.1	NA
01030300	Trout Brook near Danforth, ME	Crest stage	10	4.29	14. 9	20.7	NA
01030400	Gulliver Brook near Monarda, ME	Crest stage	11	11.3	19.9	12.8	28.0
01031470	Brewster Brook near Parkman, ME	Crest stage	14	0.65	2.89	23.1	NA
01031600	Morrison Brook near Sebec Corners, ME	Continuous	13	4.10	9.62	24.7	22.0
01034900	Coffin Brook near Lee, ME	Crest stage	11	2.04	15.3	23.2	13.7
01036380	Unnamed Brook near Bradley, ME	Crest stage	14	0.66	10.1	12.3	9.6
01037200	Shaw Brook near Northern ME Junction, ME	Crest stage	11	2.93	4.13	14.5	NA
01037430	Goose River at Rockport, ME	Crest stage	11	8.24	5.87	43.0	NA
01041900	Mountain Brook near Lake Parlin, ME	Crest stage	11	3.99	2.81	55.5	NA
01047860	Unnamed Brook near New Sharon, ME	Crest stage	14	0.58	4.64	10.3	11.0
01048840[1]	Unnamed Brook near Newport, ME	Crest stage	14	0.42	7.13	7.5	6.8
01049100	Hall Brook at Thorndike, ME	Crest stage	11	4.78	3.22	24.2	24.0
01049130	Johnson Brook at South Albion,ME	Continuous	11	2.42	15.4	16.5	10.6
01049180	Unnamed Brook near Albion, ME	Crest stage	12	0.31	2.25	23.7	9.0
01049300	North Branch Tanning Brook near Manchester, ME	Continuous	20	1.14	7.65	28.3	8.8
01049690[1]	Unnamed Brook near Dresden, ME	Crest stage	14	0.44	2.25	29.3	12.0
01049700	Gardiner Pond Brook at Dresden Mills, ME	Crest stage	10	8.02	20.6	23.4	NA
01050490	Unnamed Brook near Rangeley, ME	Crest stage	13	0.44	0.00	53.9	13.0
01050900	Four Ponds Brook near Houghton, ME	Crest stage	11	3.23	20.2	36.4	18.3
01054135	Unnamed Brook near Gilead, ME	Crest stage	14	0.34	0.02	50.6	7.4
01055300[1]	Bog Brook near Buckfield, ME	Crest stage	11	10.5	13.1	29.5	NA
01062700	Patte Brook near Bethel, ME	Crest stage	10	3.78	5.68	56.7	17.8
01063310	Stony Brook at East Sebago, ME	Continuous	18	1.55	18.5	26.2	8.6
01064200	Mill Brook near Old Orchard Beach, ME	Crest stage	10	2.30	5.13	15.1	NA
01066100	Pease Brook near Cornish, ME	Crest stage	11	5.01	3.54	49.7	17.3
01069700	Branch Brook near Kennebunk, ME	Crest stage	18	9.65	9.15	19.1	19.5

[1]These four sites were identified as having high influence in the peak-flow regression analyses (see section "Ordinary Least-Squares Regression With Bankfull Width as a Field-Based Variable").
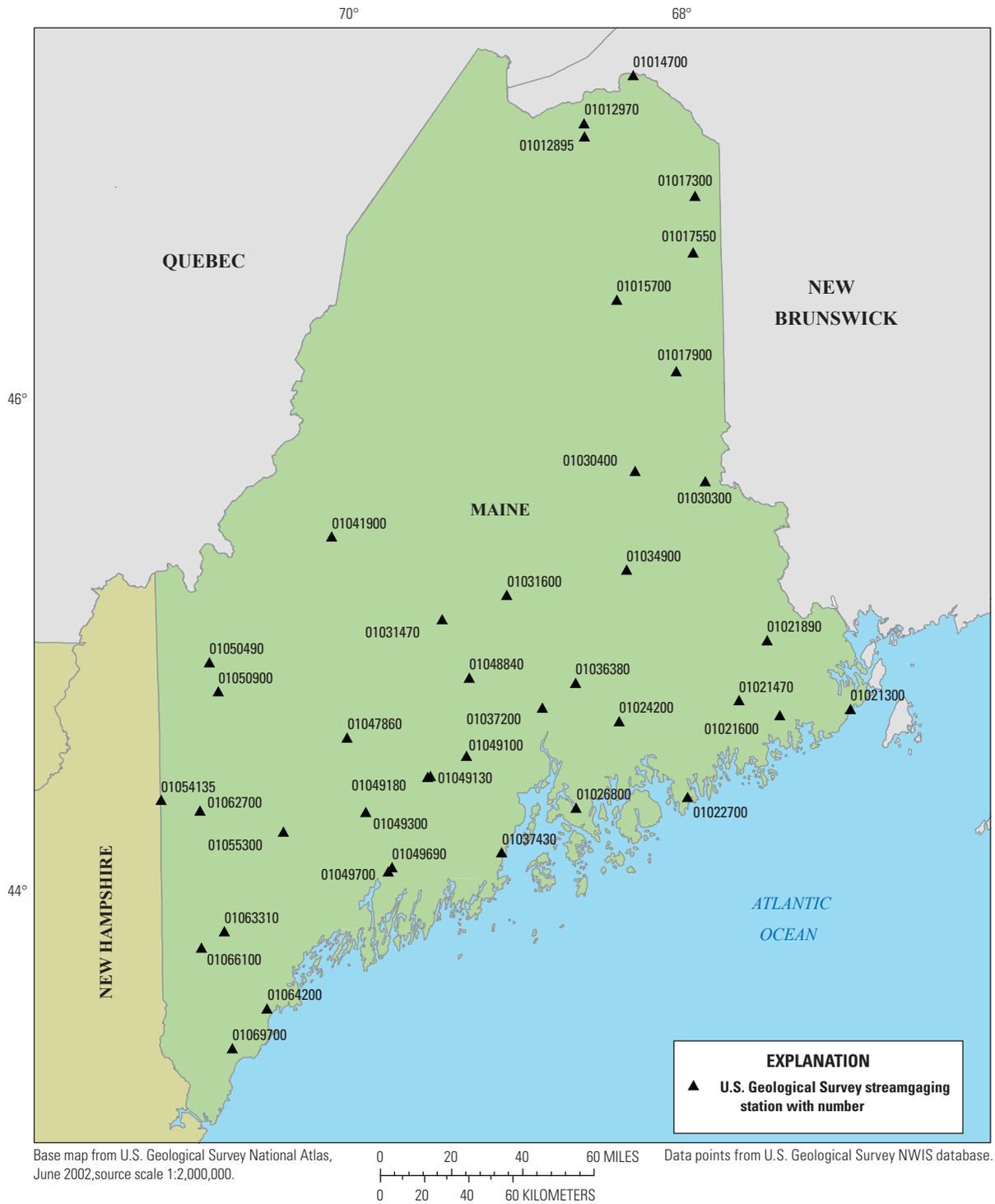
Figure 1.   Locations of the study area and 40 USGS streamgaging stations with small drainage basins in Maine used in this study.

Figure 2. Example of channel with measurable geomorphic indicators of bankfull width at Patte Brook near Bethel, Maine.

Figure 3. Example of a culvert rust line at Clark Brook near Wallagrass Plantation, Maine.

operates within ArcGIS to support geospatial and temporal data analyses. For 16 of the sites for which lidar data were not available, a digitized stereoscope method was used with Arc-Map to georeference the scanned orthophotos or topographic maps on which basin delineations were drawn. Once georeferenced, the delineations were digitized into polygons by tracing the drawn delineations on the scanned photographs. Twelve sites remained for which there were no available lidar data and no available orthophotos suitable for stereoscoping. For these sites, GIS with 10 meter (m) DEMs was used to manually delineate the drainage basins. The 28 sites for which lidar and (or) ortho stereoscoping methods were used ranged in size from 0.3 to 8.5 mi^2. The 12 sites for which only 10 m DEMs were available ranged in size from 1.5 to 11.3 mi^2.

Because manual delineation using 10 m DEM was available for all sites, the accuracy of the 10 m DEM method could be determined as compared to the accuracy of the lidar and (or) stereoscoping. Of the 28 sites for which either lidar or stereoscoping methods were used with 10 m DEM delineations, 22 had differences between the methods of less than 10 percent, 5 had differences of 10–35 percent, and 1 had a difference greater than 35 percent. Basin characteristics of average basin slope and percent basin wetlands were tested for use in the peak-flow regression equations. Percent basin wetlands was selected for use in the equations based on its relation to peak flows demonstrated in statewide peak-flow equations in Maine (Hodgkins, 1999). Average basin slope was selected as a potential variable based on its importance in the Manning's formula for calculating flows (Maidment, 1993) and the hypothesis that slope may be a better explanatory variable than basin wetlands for small, steep basins that tend to have little wetland area. Both variables were calculated for all 40 basins in the study (table 1). To determine average basin slope, the slope tool in ArcGIS was used; this tool calculates the maximum rate of change in value from neighboring cells for each grid cell. This means that the steepest gradient is identified between a cell and its neighbors, and the cell is assigned a slope value of the rise over run. Note that this method computes slope over the entire basin rather than slope of the flowline only. The GIS slope grid used in this study was produced from the National Elevation Dataset (NED) (Gesch and others, 2002) based on the best available data as of March 2014. The slope grid units were feet per feet multiplied by 100. The data used for this method were mostly based on 10 m resolution grids with sections of NED-approved lidar. Average basin slopes ranged from 7.5 to 56.7 percent for sites in the study (table 1).

Percent basin wetlands was calculated as the percentage of wetland cells in each drainage basin by using a grid derived from the National Wetland Inventory (NWI) dataset, which was produced by the U.S. Fish and Wildlife Service at a scale of 1:24,000 (U.S. Fish and Wildlife, 2008). All categories of wetlands in the NWI, including ponds and lakes, were included in the basin-wetland calculations. Percent basin wetlands ranged from 0 to 22.2 percent for sites in this study (table 1).

Peak Flow Hydrology

Statistical methods were used to compute peak flows for selected recurrence intervals at streamflow gages for use in the regression equations. Eight of the 40 streamflow gages used in this study have continuous streamflow records with 11 to 20 years of data collected between 1963 and 2012, and the remaining crest-stage gages (for which only peak flows are measured) have from 10 to 18 recorded annual peak streamflows between 1963 and 2012 (U.S. Geological Survey, 2013; table 1). Discharges for the 1-, 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year floods (equivalent to the 99-, 50-, 20-, 10-, 4-, 2-, 1-, 0.5-, and 0.2-percent AEPs, respectively) were computed at the 40 sites. The at-site estimates for all of the streamgages were determined by the standard log-Pearson type III method described in Bulletin 17B of the Hydrology Subcommittee (Interagency Advisory Committee on Water Data, 1982) and a modification of this method called the expected moments algorithm (EMA) (Cohn and others, 1997, 2001; Griffis and others, 2004) based on 10 to 20 years of peak flows. Flow frequencies were run with a generalized skew of 0.029 as developed for Maine in Hodgkins (1999). The flood-frequency calculations were performed using the U.S. Geological Survey software PeakFQ (U.S. Geological Survey, 2014).

Hydraulic Model Analysis Using a Field-Based Rust Line Variable

Hydraulic models for the seven culverts with definable rust lines were prepared for this study using the step-backwater model, Hydrologic Engineering Centers River Analysis System (HEC-RAS), version 4.1.0 (U.S. Army Corps of Engineers, 2010), in order to estimate streamflows consistent with measured rust lines. The HEC-RAS analyses were done using the steady-state flow computation option. Subcritical (tranquil) flow regime was assumed for the simulations. Flows were modeled in the channel upstream from the culvert, in the culvert, and in the channel downstream from the culvert. Starting water-surface elevations were set as the normal depth based on downstream slope estimated at each of the sites. Hydraulic analyses require the estimation of energy losses that result from frictional resistance exerted by a channel on streamflow. These energy losses are quantified by the Manning's roughness coefficient ("n" value) and were selected on the basis of field observations and high-resolution aerial photographs. The final "n" values ranged from 0.04 to 0.05 for the main channels, 0.07 to 0.08 for the overbank areas, and 0.019 to 0.034 within the culverts. Entrance losses for the culverts ranged from 0.7 to 0.9, and exit losses were typically 1. These are standard values as suggested in the HEC-RAS Hydraulic Reference Manual (U.S. Army Corps of Engineers, 2010). The HEC-RAS models were run with peak flows for selected recurrence intervals. In addition, flows were determined that best matched measured rust lines.

Of the seven culverts with definable rust lines that were surveyed and modeled, four were calibrated using both the upstream and downstream rust lines. The flows that formed the rust lines at these four culverts did not appear to be affected by downstream backwater. Flood-frequency analyses and hydraulic modeling indicated that the rust-forming flow at one of these culverts (station 01012895) corresponds to a flow with a recurrence interval slightly greater than 1 year. Rust lines at two of the culverts (stations 01021300 and 01034900) were likely formed by flows with recurrence intervals slightly less than 1 year. The rust line at the remaining culvert (station 01037430) was most likely formed by a flow with a recurrence interval much less than 1 year (table 2).

The downstream rust lines on the remaining three culverts were clearly formed by backwater conditions. The rust lines on the downstream ends were higher than would be expected due to culvert slope and friction losses and can most easily be explained by backwater caused by ice conditions, downstream channel constrictions, or temporary obstructions. This was especially evident at Hall Brook at Thorndike (USGS station number 01049100), where there was a drop at the downstream end of the culvert indicating an inlet control, but where the rust lines had likely been formed by backwater. The four sites with backwater-free rust lines were not a sufficient number for testing the elevation of the rust lines or the area beneath rust lines as explanatory variables in a peak-flow regression equation. Furthermore, the low percentage of available culverts with definable rust lines indicates that this variable is not useful for computing peak flows at ungaged basins in Maine. Culvert rust lines were thus dropped from further analyses in this study.

Regression Equations

Regression equations were created for estimating peak flows with 1-, 2-, 5-, 10-, 25-, 50-, 100-, 200-, and 500-year recurrence intervals at small (less than 12 mi^2) ungaged basins

in Maine. A limited set of potential explanatory variables was tested for use in the regression equations using ordinary least-squares (OLS) linear regression in the R statistical package (R core team, 2014). These variables included drainage area, average basin slope, percent basin wetlands developed using a GIS, and median bankfull width, which was measured in the field. The best explanatory variables for the regression equations were determined by choosing variables that resulted in the lowest regression errors. If errors were similar, variables that were easier to compute and ones that best explained peak-flow variability across the entire range of recurrence intervals (from 1 to 500 years) were favored. After choosing explanatory variables with OLS, final equation coefficients were computed using generalized least-squares (GLS) regression analyses (Stedinger and Tasker, 1985) using the Weighted-Multiple-Linear Regression (WREG) program (Eng and others, 2009).

Ordinary Least-Squares Regression With Bankfull Width as a Field-Based Variable

Only 23 of the 40 sites had usable measurements of bankfull width; thus, initial OLS regression analyses were done using this subset of 23 sites with combinations of drainage area, percent basin wetlands, average basin slope, and bankfull width as explanatory variables and flows with 1- to 500-year recurrence intervals as response variables. Regression diagnostics during exploratory analyses included residual plots that checked for linearity, homoscedasicity (constant variance), normality, and the presence of outliers. Residual plots indicated whether explanatory variables showed bias over their ranges and whether variables needed to be transformed to ensure a linear relation between explanatory and response variables, a requirement for satisfactory regressions. A base-10 logarithmic transformation of the drainage-area explanatory variable and all peak-flow response variables was tested and adopted because this transformation

Table 2.   Rust-forming flows at seven culverts in Maine.

[U.S. Geological Survey; ME, Maine; ft^3/s, cubic feet per second]

USGS streamgage number	Streamgage name	Flow with a 1-year recurrence interval, (ft^3/s)	Rust-forming flow, (ft^3/s)
01012895	Clark Brook near Wallagrass Plantation, ME (Eagle Lake)	10.6	15.0
01021300	Wiggins Brook near West Lubec, ME	88.5	67.0
01021600	Middle River near Machias, ME	100	Rust line influenced by backwater
01024200	Garland Brook near Mariaville, ME	119	Rust line influenced by backwater
01034900	Coffin Brook near Lee, ME	21.8	17.0
01037430	Goose River at Rockport, ME	135	28.0
01049100	Hall Brook at Thorndike, ME	37.1	Rust line influenced by backwater

improved the fit of the regression equations. Multicollinearity among the explanatory variables (variables that are highly correlated) was measured using the variance inflation factor, and no problems with multicollinearity were noted between the final variables.

Influence of individual stations on the regression equations was measured using the Cook's D statistic. High influence means that a site is both an outlier and has high leverage. Although the Cook's D statistic indicated that four sites exerted high influence on the final regression equations (table 1), these sites were included in the final analyses because there were no obvious problems with the data that would warrant their removal. Alternatively, if these four sites had a common basin characteristic, all sites with that basin characteristic could be removed and the equations would not apply to any sites with this same basin characteristic—or the common basin characteristic could indicate the need for additional explanatory variables. No common basin characteristic was found for these four sites.

The 23-site model with drainage area and percent basin wetlands explained more variability in the peak flows than did drainage area and average basin slope, drainage area by itself, or bankfull width by itself as based on the R^2 values of the models. The model with drainage area and percent basin wetlands explained 75 to 84 percent of the variability for peak flows with 1- to 500-year recurrence intervals which was slightly better than the model that included drainage area combined with bankfull width that explained 73 to 83 percent of the variability for these peak flows. Since bankfull width is more time consuming to calculate (involving a field visit) and is not available for all sites, bankfull width was removed from further analyses.

Ordinary least-squares regressions were computed again using the remaining variables for all 40 sites. Once again, the model with drainage area and percent basin wetlands explained the most variability; explaining 75 to 84 percent of the variability in peak flows with 1- to 500-year recurrence intervals. Drainage area by itself explained almost as much variability as drainage area combined with percent basin wetlands for the 1-year recurrence interval but much less of the variability for the 500-year recurrence interval, and thus basin wetlands was retained in all of the models for consistency. The model with slope and drainage area explained 68 to 79 percent of the variability in peak flows with 1- to 500-year recurrence intervals and thus did not perform as well as the model with wetlands and drainage area.

Generalized Least-Squares Regression Analyses

Final equation coefficients and accuracy were computed with GLS regression analyses using the WREG program (Eng and others, 2009) and are presented in table 3. The equations are of the form

$$Q_{1...500} = bA^a10^{-w(W)}$$

where Q_1 through Q_{500} are discharges with 1- to 500-year recurrence intervals, A is drainage area, W is basin wetlands, and the coefficients b, a, and w vary with recurrence interval.

Stedinger and Tasker (1985) found that GLS regression equations are more accurate and provide better estimates of uncertainty than OLS regression equations when streamflow records are of different and widely varying lengths and when concurrent peak flows are correlated. Concurrent peak flows across Maine can be correlated because rainstorms can affect large areas and snowmelt-related runoff tends to be regionally coherent. Less weight is given to GLS regression techniques used at stations where the streamflow record is more highly correlated with the concurrent record of other stations.

Equation accuracy is quantified by the average standard error of prediction (ASEP), which is a measure of how well the regression equations will estimate the flow statistic of interest when applied to an ungaged, unregulated, rural basin in Maine. The probability that the true value of the flow statistic at a site is between the negative- and positive-percent ASEP is approximately 68 percent. For example, there is a 68-percent probability that the true peak-flow value with a 100-year recurrence interval is between –33.3 and 50.0 percent of the 100-year peak flow value computed with the regression equation (table 3).

Limitations of Regression Equations

When estimating peak flows for ungaged basins in Maine on the basis of the regression equations described in this

Table 3.   Regression equations and their accuracy for estimating peak streamflows for small, ungaged streams in Maine that are rural and unregulated.

[ASEP, average standard error of prediction; Q_1 through Q_{500}, peak flows with 1- to 500-year recurrence intervals in cubic feet per second; A, drainage area in square miles; W, basin wetlands in percent]

Regression equation	ASEP (in percent)
$Q_1 = 18.836(A)^{.773}10^{-.013(W)}$	-31.0 to 45.0
$Q_2 = 64.121(A)^{.803}10^{-.015(W)}$	-28.4 to 39.6
$Q_5 = 102.565(A)^{.809}10^{-.016(W)}$	-28.9 to 40.7
$Q_{10} = 131.522(A)^{.811}10^{-.017(W)}$	-30.0 to 42.9
$Q_{25} = 171.791(A)^{.814}10^{-.017(W)}$	-31.5 to 46.0
$Q_{50} = 204.174(A)^{.816}10^{-.018(W)}$	-32.4 to 48.0
$Q_{100} = 238.781(A)^{.817}10^{-.018(W)}$	-33.3 to 50.0
$Q_{250} = 275.423(A)^{.818}10^{-.019(W)}$	-34.6 to 52.9
$Q_{500} = 327.341(A)^{.820}10^{-.019(W)}$	-35.8 to 55.7

report, it is important that the explanatory variables of drainage area and percent basin wetlands be derived using the same or comparable methods as those documented in this report. If basin characteristics are derived using different methods, estimates of flow computed from the equations will be of unknown accuracy. If the values of drainage area and percent basin wetlands at a site are outside the two-dimensional range of those used to develop these equations (fig. 4, outside shaded area), estimates will also be of unknown accuracy. The equations in this report were developed using rural basins in Maine and thus should be applied to rural basins in Maine. Hodgkins (1999) contains guidance on estimating peak flows for urbanized basins. There were no peak-flow data available for small basins in the northwestern part of Maine (fig. 1); therefore, the equations may not apply to that area of the State.

Basins included in this study had no substantial peak-flow reservoir regulation; therefore, estimates of peak flows at small, ungaged basins, using the equations in this report, are unlikely to be accurate for basins with substantial peak-flow regulation. Benson (1962) defines substantial reservoir regulation as more than 4.5 million cubic feet per square mile (or approximately 100 acre-feet per square mile) of usable reservoir storage. Usable storage is the volume of water available for release from a reservoir, between the minimum and maximum controllable elevations.

Evaluation of Regression Equations

A previous peak-flow study for Maine developed regression equations to estimate peak flows on basins with drainage areas of 1 to 1,700 mi^2 (Hodgkins, 1999). The final explanatory variables in Hodgkins (1999) were drainage area and percent basin wetlands, the same as in the current report; however, many more potential explanatory variables were tested in the 1999 report. Basin slope was tested in both equations, but wetlands combined with drainage area was found to be a better predictor of peak flows than slope combined with drainage area in both reports.

The area of overlap between the 1999 equations and the current equations includes basins with drainage areas between 1 and 12 mi^2; both sets of equations were designed to be used within this range. The equations can result in very different estimates of peak flows in this range due to the different sets of stations from which they were developed. Differences between estimates from the two sets of equations are most pronounced for estimates of high recurrence-interval peak flows for basins with a low percentage of wetlands (figs. 5 and 6).

Thirty sites have drainage areas between 1 and 12 mi^2 and sufficient streamflow data for testing the accuracy of the equations. For each of the 30 basins, differences between the predicted peaks from the two sets of regression equations were compared to observed peaks. Overall errors between predicted and observed peaks were computed with the root mean square error (RMSE). Analyses of RMSEs indicate that errors are less for the equations in the current study as opposed to equivalent equations in the 1999 study. For example, RMSEs for peak flows with a 10-year recurrence interval are –30.9 to 44.8 percent for the 1999 equations, but are –27.4 to 37.7 percent for the current equations (table 4). Thus, the equations in the current study should be used for ungaged basins in Maine with drainage areas from 1 to 12 mi^2 because these equations are expected to provide estimates that are closer to the true values for these small basins.
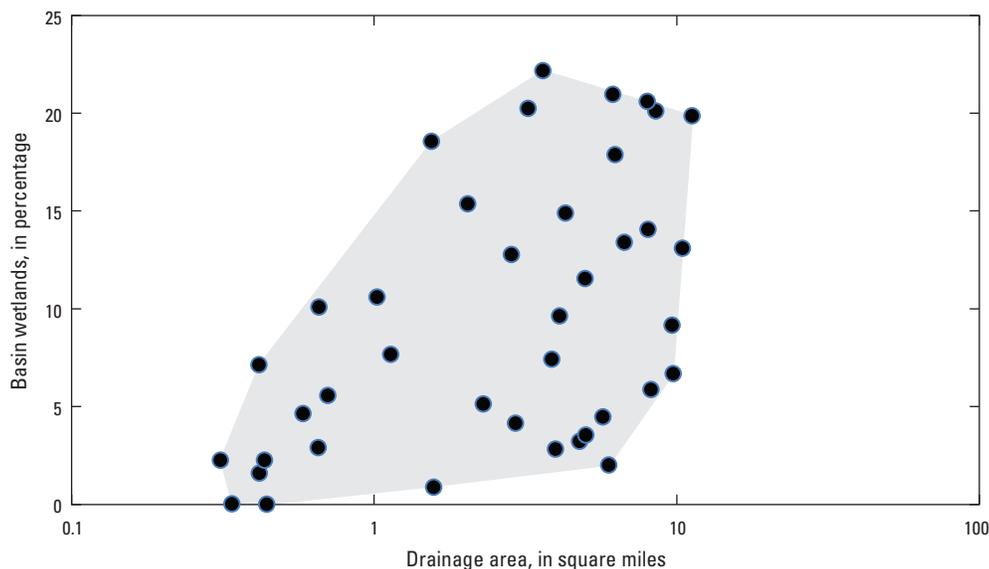

Figure 4.   Two-dimensional range of explanatory variables for the regression equations.

5 percent basin wetlands

Figure 5. Differences in peak flows with 2-year and 100-year recurrence intervals, based on Hodgkins (1999) equations and equations from the current study, for drainage areas between 1 and 10 square miles and holding percent basin wetlands constant at 5 percent.
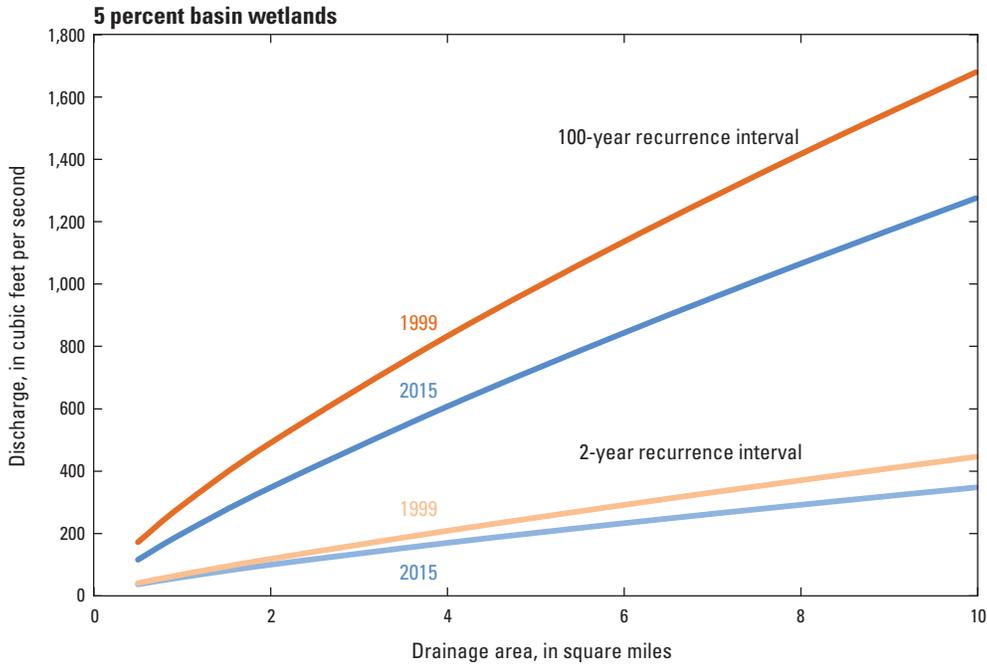
15 percent basin wetlands

Figure 6. Differences in peak flows with 2-year and 100-year recurrence intervals, based on Hodgkins (1999) equations and equations from the current study, for drainage areas between 1 and 10 square miles and holding percent basin wetlands constant at 15 percent.
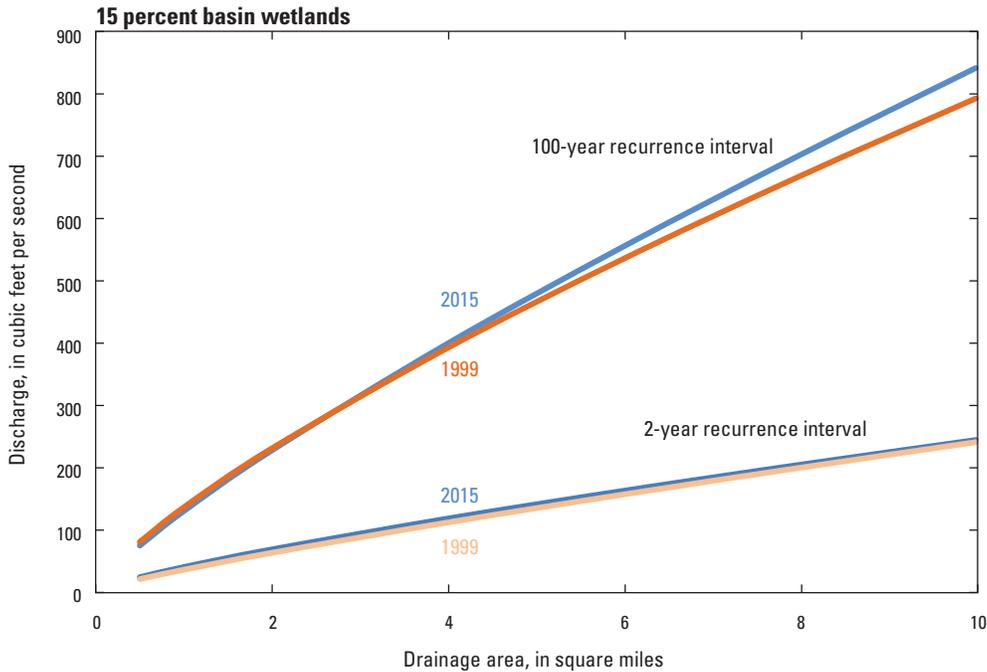
Table 4. Comparison of errors of the 1999 peak-flow equations and errors of the equations from the current study when each set of equations is applied to basins with drainage areas between 1 and 12 square miles.

[RMSE, root mean square error]

Recurrence interval	1999 equations RMSE (in percent)	Current study equations RMSE (in percent)
2 years	-27.5 to 37.9	-23.1 to 30.1
5 years	-29.3 to 41.4	-25.3 to 33.8
10 years	-30.9 to 44.8	-27.4 to 37.7
25 years	-33.0 to 49.2	-29.9 to 42.6
50 years	-34.5 to 52.7	-31.9 to 46.8
100 years	-35.9 to 56.1	-33.6 to 50.7
500 years	-39.1 to 64.1	-37.6 to 60.2

Comparisons of RMSE errors between the two sets of equations were also examined for the 20 stations with the smallest drainage areas in Maine that have sufficient data. These stations have drainage areas between 0.3 and 3.6 mi^2. Although the 1999 equations were not designed to be used on basins less than 1 mi^2 and because data were not available in this range at the time they were developed, the equations can now be tested in this range with the newer data. For these 20 small basins, the current equations perform better than do the 1999 equations. For example, RMSEs for peak flows with a 500-year recurrence interval are –50.3 to 101.1 percent for the 1999 equations, but are –37.6 to 60.3 percent for the current equations (table 5). The equations developed by Hodgkins (1999) should continue to be used for all basin areas greater than 12 mi^2.

Table 5. Comparison of errors of the 1999 peak-flow equations and errors of the equations from the current study when each set of equations is applied to basins with drainage areas between 0.3 and 3.6 square miles.

[RMSE, root mean square error]

Recurrence interval	1999 equations RMSE (in percent)	Current study equations RMSE (in percent)
2 years	-33.7 to 50.9	-30.7 to 44.2
5 years	-37.4 to 59.7	-31.1 to 45.2
10 years	-40.0 to 66.6	-32.1 to 47.2
25 years	-43.1 to 75.6	-33.2 to 49.7
50 years	-45.0 to 81.9	-34.4 to 52.3
100 years	-46.8 to 88.1	-35.3 to 54.5
500 years	-50.3 to 101.1	-37.6 to 60.3

Summary

In 2011, the U.S. Geological Survey began a cooperative investigation with the Maine Department of Transportation (MaineDOT) to test regression-equation explanatory variables for estimating peak flows on small (less than 12 square miles [mi^2]) basins in Maine that have 1- to 500-year recurrence intervals. These recurrence intervals are equivalent to floods that have a 99- to 0.2-percent chance of being exceeded each year. Peak-flow data from 40 USGS streamgages that had small drainage basins and a minimum of 10 years of recorded peak flows were used in this study. A major objective of the study was to determine if field-based characteristics, such as culvert rust lines or bankfull channel widths, could improve peak-flow regression equations. Seven of the 40 sites had well-defined rust lines in a culvert near a gage and three of the seven had rust lines that were likely created by downstream backwater; thus, the rust line variable was determined not to be useful for estimating peak flows in Maine. Twenty-three of the 40 gages had definable bankfull widths near gages. Bankfull width was tested as an explanatory variable in the regression equations, but it did not improve the equations over easier-to-compute variables such as percent basin wetlands in the basin and, thus, was not selected as a variable in the final peak-flow equations.

Ordinary least-squares regression analyses were used to determine the variables that best explained peak-flow statistics and to examine regression equation diagnostics to ensure that regression assumptions were being met. The best variables were basin area and percent basin wetlands. Generalized least-squares regression analyses using 40 gages were used to derive the final regression coefficients for drainage area and percent basin wetlands and the measures of uncertainty for the regression equations.

Regression equations presented here are the best available equations for estimating peak flows on ungaged rural, unregulated basins in Maine with drainage areas of 0.3 to 12 mi^2 and percent basin wetlands of 0 to 22.2 percent of the drainage area. Equations developed by Hodgkins (1999) continue to be the best available equations for estimating peak flows for ungaged basins with areas greater than 12 mi^2. Hodgkins (1999) also contains guidance on estimating peak flows for urbanized basins. Drainage areas and percent basin wetland basin characteristics should be derived using the same or comparable methods as those documented in this report. Using basin characteristics outside the range of those used to derive the equations, or using different methods of computing the basin characteristics, will lead to estimates of unknown accuracy.

References Cited

Benson, M.A., 1962, Factors influencing the occurrence of floods in a humid region of diverse terrain: U.S. Geological Survey Water-Supply Paper 1580–B, 64 p.

Cohn, T.A., Lane, W.M., and Baier, W.G., 1997, An algorithm for computing moments-based flood quantile estimates when historical flood information is available: Water Resources Research, v. 33, no. 9, p. 2089–2096.

Cohn, T.A., Lane, W.M., and Stedinger, J.R., 2001, Confidence intervals for expected moments algorithm flood quantile estimates: Water Resources Research, v. 37, no. 6, p. 1695–1706.

Dudley, R.W., 2004, Estimating monthly, annual, and low 7-day, 10-year streamflows for ungaged rivers in Maine: U.S. Geological Survey Scientific Investigations Report 2004–5026, 22 p.

Dunne, T., and Leopold, L.B., 1978, Water in environmental planning: San Francisco, Calif., Freeman and Co., 818 p.

Eng, Ken, Chen, Yin-Yu, and Kiang, J.E., 2009, User's guide to the weighted-multiple-linear-regression program (WREG version 1.0): U.S. Geological Survey Techniques and Methods, book 4, chap. A8, 21 p. (Also available at http://pubs.usgs.gov/tm/tm4a8/.)

Gesch, D., Oimoen, M., Greenlee, S., Nelson, C., Steuck, M., and Tyler, D., 2002, The National Elevation Dataset: Photogrammetric Engineering and Remote Sensing, v. 68, no. 1, p. 5–11.

Griffis, V.W., Stedinger, J.R., and Cohn, T.A., 2004, Log Pearson type 3 quantile estimators with regional skew information and low outlier adjustments: Water Resources Research, v. 40, no. 7, W07503, 17 p., accessed July 7, 2014, at doi:10.1029/2003WR002697.

Hodgkins, G.A., 1999, Estimating the magnitude of peak flows for streams in Maine for selected recurrence intervals: U.S. Geological Survey Water-Resources Investigations Report 99–4008, 45 p.

Hodgkins, G.A., Hebson, Charles, Lombard, P.J., and Mann, Alexander, 2007, Comparison of peak-flow estimation methods for small drainage basins in Maine: U.S. Geological Survey Scientific Investigations Report 2007–5170, 32 p.

Interagency Advisory Committee on Water Data, 1982, Guidelines for determining flood flow frequency: U.S. Geological Survey, Bulletin 17B of the Hydrology Subcommittee, Office of Water Data Coordination, 183 p., accessed May 12, 2014, at http://water.usgs.gov/osw/bulletin17b/dl_flow.pdf.

Leopold, L.B., and Maddock, T., Jr., 1953, The hydraulic-geometry of river channels and some physiographic implications: U.S. Geological Survey Professional Paper 252, 57 p.

Lombard, P.J., 2004, August median streamflow on ungaged streams in eastern coastal Maine: U.S. Geological Survey Scientific Investigations Report 2004–5157, 15 p.

Lombard, P.J., 2010, June and August median streamflows estimated for ungaged streams in southern Maine: U.S. Geological Survey Scientific Investigations Report 2010–5179, 16 p.

Lombard, P.J., Tasker, G.D., and Nielsen, M.G., 2003, August median streamflow on ungaged streams in eastern Aroostook County, Maine: U.S. Geological Survey Water-Resources Investigations Report 03–4225, 19 p.

Maidment, D.R., 1993, Handbook of hydrology: New York, McGraw-Hill, 1,424 p.

R Core Team, 2014, The R project for statistical computing: Vienna Austria, R Foundation for Statistical Computing, accessed March 6, 2015, at http://www.r-project.org/.

Stedinger, J.R., and Tasker, G.D., 1985, Regional hydrologic analysis 1, Ordinary, weighted, and generalized least-squares compared: Water Resources Research, v. 21, no. 9, p. 1421–1432.

U.S. Army Corps of Engineers, Hydrologic Engineering Center, 2010, HEC-RAS River Analysis System, Hydraulic Reference Manual, version 4.1.0, [variously paged].

U.S. Fish and Wildlife Service, 2008, National Wetlands Inventory Maps, vector digital data from Maine Office of Geographic Information Systems, scale 1:24,000, accessed March 2008, at http://megis.maine.gov/catalog/.

U.S. Geological Survey, 2013, Water resources application software, PeakFQ, accessed 2013, at http://water.usgs.gov/software/PeakFQ/.

U.S. Geological Survey, 2014, USGS surface-water data for Maine: National Water Information System, accessed August 2014, at: http://waterdata.usgs.gov/me/nwis/sw/.