Go with the Flow

Anna Baumgartner
Anupama Rao
Eiluned Roberts
University of Alaska—Fairbanks
Fairbanks, AK 99775

Advisor: Robert A. Hollister

Introduction

We develop a method of estimating the flow rate from a water tank and the total amount of water used during a specified time period, given any time scale and water-level measurements taken at nearly regular intervals. Our analysis includes the situation where the tank is refilled at an unknown rate while water is flowing from it.

The problem can be divided into three major steps:

- determine numerically the flow rate at the given data points;
- find a smooth-fitting approximation of the flow rate out of the tank, and
- take the antiderivative of the approximation to the flow rate to cancel the pump's contribution to the original data.

By following these steps, we approximate the flow rate, using the original data and a cubic-spline fit. We also reconstruct data during the operation of the refill pump, in order to calculate the total volume of water consumed in a 24-hour period.

Assumptions and Justifications

- The given data have measurement errors of no more than 0.5%, as specified in the problem statement.
- The day arbitrarily starts at time $t = 0$ seconds and ends at $t = 86,400$ seconds. If we start the day at a different time, the flow rate function remains the same but the total volume of water consumed changes slightly.
- The state water-rights agencies do not require instantaneous flow rates but rather are interested in the general trends in water use over the day. Features of the data, such as the peak hours of water use, would be more appropriate for such an agency.

- The flow rate can be approximated by a smooth curve. That is, the second derivative of the flow rate is continuous at the points given.

Developing the Model

Finding the Approximate Flow Rate

Our greatest difficulty is to describe the behavior of the flow rate during replenishment of the tank. To overcome this problem, we estimate the first derivative (the flow rate) at each given data point, excluding those when the pump was in operation.

Since our data closely resemble evenly-spaced data points, we use a slightly modified form of first-derivative formulas from interpolating polynomials for data of such type [Gerald et al. 1989]. We divide the data points into the three groups shown in Figure 1.

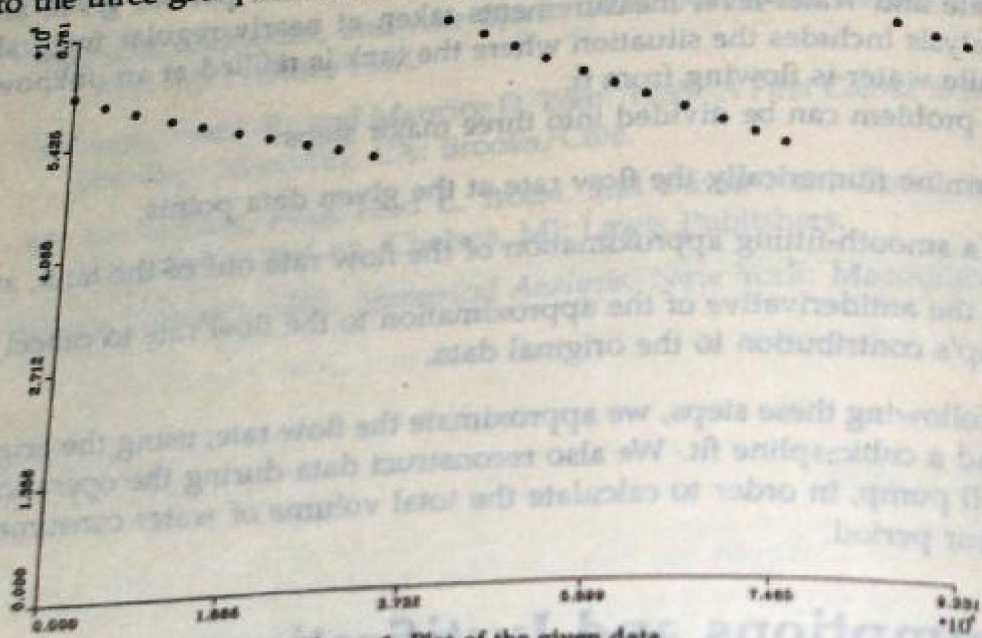

Figure 1. Plot of the given data.

Let $y = f(t)$ be the volume of water in the tank at time t. We use the following formula, derived from the well-known *Newton-Gregory forward polynomial*, for the first two points in each group:

$$f'(t_i) \doteq \sum_{j=1}^{n} \frac{(-1)^{j+1} \Delta f_{i+j}}{j h_{i+j}},$$

where (t_i, f_i) is the data point in question, (t_{i+1}, f_{i+1}) is the next data point, Δf_i is a forward difference, e.g., $(f_{i+1} - f_i)$, and similarly, $h_i = t_{i+1} - t_i$. To simplify computations, we choose $n = 3$.

For the last two points of each group, we use the Newton-Gregory *backward* polynomial:

$$f'(t_i) \doteq \frac{\Delta f_{i-1}}{h_{i-1}} + \sum_{j=n}^{-2} \frac{(-1)^{j+1}\Delta f_{i+j}}{j h_{i+j}}.$$

Note that (t_{i-1}, f_{i-1}) precedes (t_i, f_i). Here we choose $n = -3$ to simplify calculations.

Finally, for intermediate data points, we use central differences, to reduce our error by using more surrounding points:

$$f'(t_i) \doteq \frac{-f_{i+2} + 8f_{i+1} - 8f_{i-1} + f_{i-2}}{12h_i}.$$

We used a computer program [EDITOR'S NOTE: Omitted.] for these numerical differentiations to get a fairly accurate approximation of the flow rate, excluding the points when the refill pump was in operation.

Interpolating the Derivative Using a Cubic Spline

Our consequent derivative approximation is very rough in the regions that contain unknown points, so we use a cubic spline to find a smooth curve to represent f', the flow rate. The cubic spline not only fits all of the points in the derivative plot, but fits them in a very pleasing way, as shown in Figure 2.

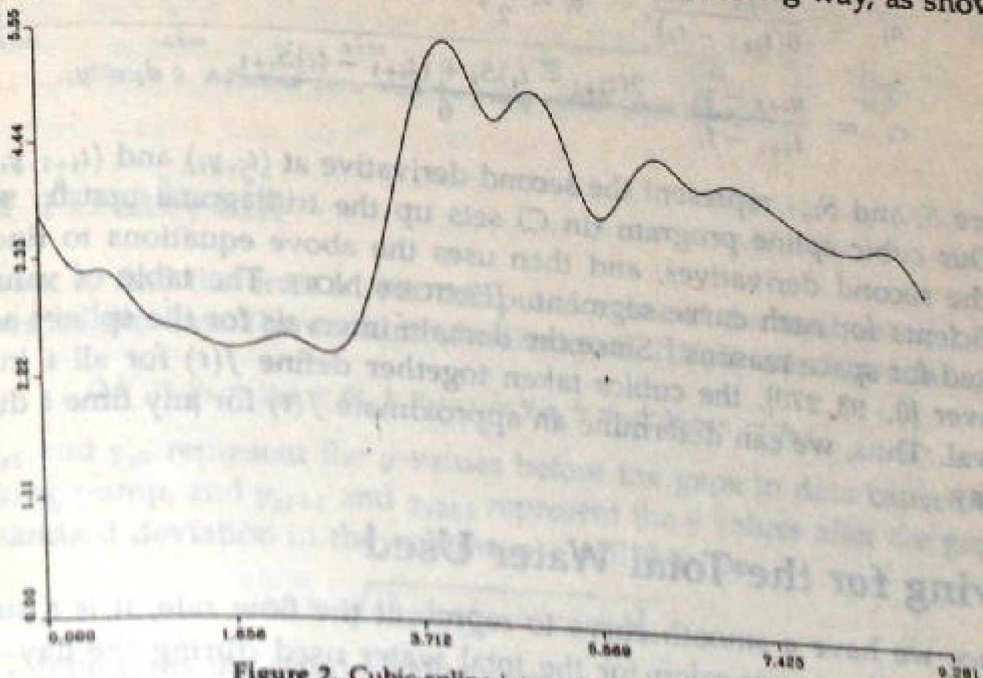

Figure 2. Cubic-spline interpolation of flow rate.

Cubic splines are well-known, so we include only a sketch of our method. For a more complete development of the cubic spline method, see Ahlberg et al. [1967].

First, we assume that the interpolated curve through n data points, with graph endpoints t_1 and t_n, is written as a series of cubic equations, one for each interval $[t_i, t_{i+1}]$. Algebraically, this assumption is expressed as

$$y = a_i(t-t_i)^3 + b_i(t-t_i)^2 + c_i(t-t_i) + d_i.$$

Second, we assume that the slope and curvature of the endpoints match up for each joining pair of cubics, to produce the smoothing effect—the basic motivation behind cubic-spline interpolation. Equivalently, first and second derivatives of the curve segments must match up at the endpoints of the cubics. Consequently, we write constraining equations

$$\begin{aligned} y_i &= a_i(t_i-t_i)^3 + b_i(t_i-t_i)^2 + c_i(t_i-t_i) + d_i \\ &= a_{i-1}(t_i-t_{i-1})^3 + b_{i-1}(t_i-t_{i-1})^2 + c_{i-1}(t_i-t_{i-1}) + d_{i-1}, \\ y_i' &= 3a_i(t_i-t_i)^2 + 2b_i(t_i-t_i) + c_i \\ &= 3a_{i-1}(t_i-t_{i-1})^2 + 2b_{i-1}(t_i-t_{i-1}) + c_{i-1}, \\ y_i'' &= 6a_i(t_i-t_i) + 2b_i = 6a_{i-1}(t_i-t_{i-1}) + 2b_{i-1}. \end{aligned}$$

We solve these equations for the coefficients a_i, b_i, c_i, and d_i by first calculating the second derivative of each function, then using the resulting tridiagonal matrix and substitution to get

$$\begin{aligned} a_i &= \frac{S_{i+1}-S_i}{6(t_{i+1}-t_i)}, \qquad b_i = \frac{S_i}{2}, \\ c_i &= \frac{y_{i+1}-y_i}{t_{i+1}-t_i} - \frac{2(t_{i+1}-t_i)S_i + (t_{i+1}-t_i)S_{i+1}}{6}, \qquad d_i = y_i, \end{aligned}$$

where S_i and S_{i+1} represent the second derivative at (t_i, y_i) and (t_{i+1}, y_{i+1}).

Our cubic spline program (in C) sets up the tridiagonal matrix, solves for the second derivatives, and then uses the above equations to find the coefficients for each curve segment. [EDITOR'S NOTE: The table of values is omitted for space reasons.] Since the domain intervals for the splines adjoin to cover [0, 93,270], the cubics taken together define $f(t)$ for all t in that interval. Thus, we can determine an approximate $f(t)$ for any time t during the day.

Solving for the Total Water Used

Since we have a smooth curve to represent the flow rate, it is a simple matter to find an expression for the total water used during the day—just integrate the flow rate over each interval when the pump was filling the tank. The integration is easy, as the equations are cubic polynomials.

We use this information to fill in the missing values in the original data and to transpose the portions of our data which have been affected by the flow of water from the pump to what we refer to as "adjusted data." [EDITOR'S NOTE: The table of values is omitted for space reasons.]. Put simply, the adjusted data are a projection of what the graph of the water volume would be if the pump had never refilled the tank. Now we can find the total amount of water used during the day by subtracting the amount of water in the tank at $t = 86{,}400$ from the amount of water at $t = 0$ (see Figure 3). Applying this method, we determine that the total amount of water used in a 24-hour period is about 326,600 gallons.

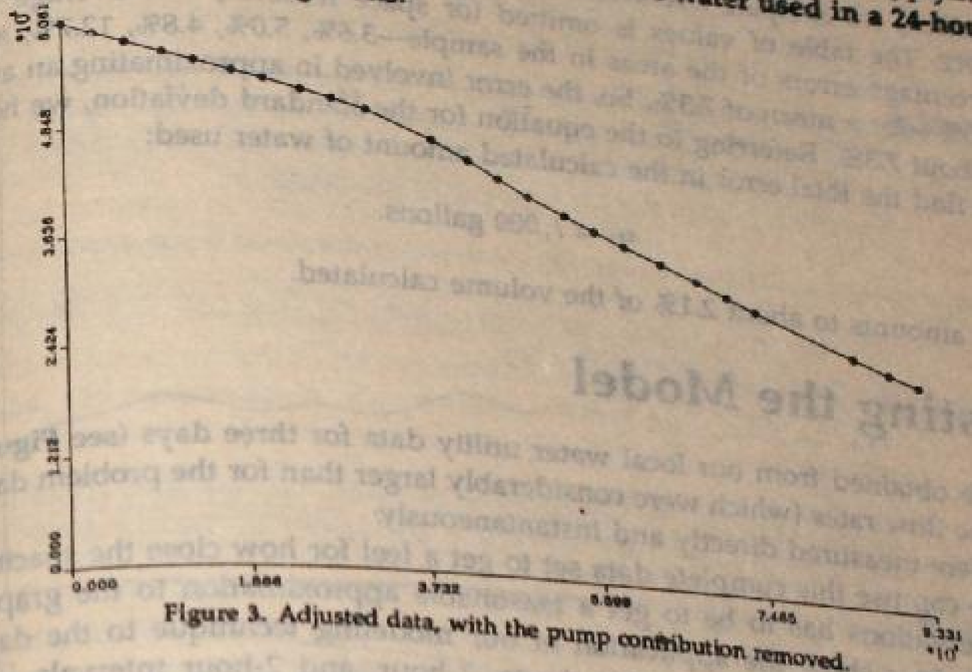

Figure 3. Adjusted data, with the pump contribution removed.

Error Analysis

We want to find the error in our computed value of the total amount of water used in one day. The amount of water used, ΔV, is calculated as

$$\Delta V = y_1 - y_{g1} + g_1 + y_{g1+1} - y_{g2} + g_2 + y_{g2+1} - y_n,$$

where y_{g1} and y_{g2} represent the y-values before the gaps in data caused by the working pump, and y_{g1+1} and y_{g2+1} represent the y values after the gaps. So the standard deviation in the volume, s_V, will then be

$$s_V = \sqrt{s_{y_1}^2 + s_{y_{g1}}^2 + \cdots + s_{y_n}^2},$$

where s_X represents the standard deviation in X [Beers 1957].

We are given the errors in the y-values from the problem statement and our corresponding assumption, namely, $s_{y_i} = 0.005y_i$.[1]

[1][EDITOR'S NOTE: It would be more customary to take the given error bound to be two standard deviations.]

To find the errors in the g-values (gap approximations), we choose random sections from areas of the graph of the flow rate which are not affected by the flow of the pump, and compare the actual area with the approximation from our cubic-spline methods. We estimate the error in the actual area to be 0.5%.

To estimate the error between an actual area and its corresponding approximation, we compare five actual areas with their approximations. [EDITOR'S NOTE: The table of values is omitted for space reasons.] We average the percentage errors of the areas in the sample—3.6%, 5.0%, 4.8%, 12.7%, and 10.3%—for a mean of 7.3%. So, the error involved in approximating an area is about 7.3%. Referring to the equation for the standard deviation, we now can find the total error in the calculated amount of water used:

$$s_V = 7,000 \text{ gallons.}$$

This amounts to about 2.1% of the volume calculated.

Testing the Model

We obtained from our local water utility data for three days (see **Figure 4**). The flow rates (which were considerably larger than for the problem data set) were measured directly and instantaneously.

We can use this complete data set to get a feel for how close the spacing of observations has to be to get a reasonable approximation to the graph. **Figures 6a–c** show the application of our modeling technique to the data set, using the data values at 30-minute, 1-hour, and 2-hour intervals. In comparing these figures to the continuous data of **Figure 2**, we note that **Figure 5b**, with data at 1-hour intervals, is a fair approximation.

Figure 4. Seventeen hours of data from the local water utility.

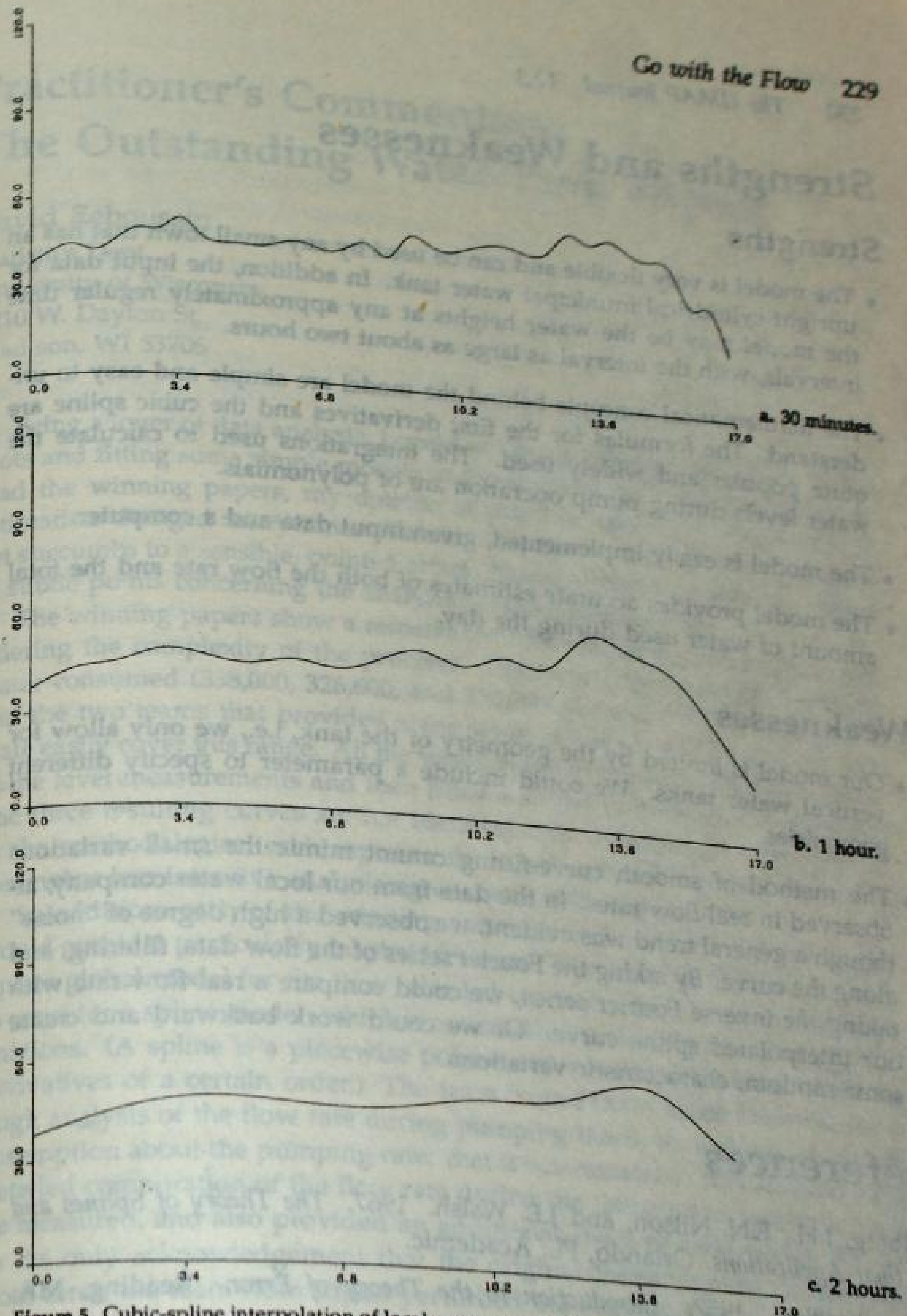

Figure 5. Cubic-spline interpolation of local water-utility data, using data at intervals of: a. 30 minutes. b. 1 hour. c. 2 hours.

Strengths and Weaknesses

Strengths

- The model is very flexible and can be used by any small town that has an upright cylindrical municipal water tank. In addition, the input data for the model may be the water heights at any approximately regular time intervals, with the interval as large as about two hours.
- The mathematical concepts behind the model are simple and easy to understand. The formulas for the first derivatives and the cubic spline are quite popular and widely used. The integrations used to calculate the water levels during pump operation are of polynomials.
- The model is easily implemented, given input data and a computer.
- The model provides accurate estimates of both the flow rate and the total amount of water used during the day.

Weaknesses

- Our model is limited by the geometry of the tank, i.e., we only allow for vertical water tanks. We could include a parameter to specify different geometries.
- The method of smooth curve-fitting cannot mimic the small variations observed in real flow rates. In the data from our local water company, although a general trend was evident, we observed a high degree of "noise" along the curve. By taking the Fourier series of the flow data, filtering, and taking the inverse Fourier series, we could compare a real flow rate with our interpolated spline curve. Or we could work backward and create some random, characteristic variations.

References

Ahlberg, J.H., E.N. Nilson, and J.L. Walsh. 1967. *The Theory of Splines and their Applications*. Orlando, FL: Academic.

Beers, Yardley. 1957. *Introduction to the Theory of Error*. Reading, MA: Addison-Wesley.

Gerald, C.F., and P.O. Wheatly. 1989. *Applied Numerical Analysis*. 4th ed. Reading, MA: Addison-Wesley.