

Generating stream rating information from data

Alternative Hydraulics Paper 8

John D. Fenton

Institute of Hydraulic and Water Resources Engineering
Vienna University of Technology, Karlsplatz 13/222,
1040 Vienna, Austria

<http://johndfenton.com/>
<mailto:JohnDFenton@gmail.com>

Wednesday 4th November, 2015

The approximation of stream rating data to generate rating curves is a field which has seen relatively little adoption of systematic numerical methods. Instead, it has been hampered by the adherence to a single power function obtained from simplifications of elementary hydraulics formulae. In general there is no reason for any part of the rating curve, even for hydraulic structures, to follow such a function. Its use has led to unnecessarily complicated, rigid, and labour-intensive manual methods for the generation of rating curves.

Alternatives are obvious, and are known. A simpler and more automatic method is just to use piecewise linear interpolation, whether on natural, logarithmic or square root axes. Cubic splines give a more continuous and aesthetically more pleasing rating curve.

A better approach than interpolation is to use least squares approximation, as it does not require the partial input of desired results at computational points. Global polynomial approximation methods are discussed, revealing some dangers that may make them fragile and unreliable, which might explain their lack of adoption by the profession. A computationally-robust method is developed which overcomes most such problems.

An even more robust method is the use of splines together with least-squares approximation. A theory and methodology for such approximating splines is developed.

Both least-squares approximation methods, global polynomial and splines, are found to work well. Both allow the specification of a weight for each data point. These weights enable the filtering of data or the incorporation of the age of data into the approximations, revealing changes in the stream and giving a rating curve dominated by the most recent data.

The nature of rating data is discussed. It is suggested that any scatter of points may be due to changes in resistance of the stream due to bed changes. They can be quite ephemeral, so that we may not know what the real rating is even shortly after gauging. If we could see the continuous stage-discharge trajectory this would be called hysteresis or "loopiness". As we usually do not, such effects appear as data scatter. They are probably always present to some degree. A method is presented for incorporating the scatter by calculating a *rating envelope* so that for routine stage measurements not only the most likely discharge could be published, but also the maximum and minimum possible values. An unintended benefit of this might be the recognition that what we are doing is not exact, and possibly not even precise.

This report is: Fenton, J. D. (2015) Generating stream rating information from data, Alternative Hydraulics Paper 8, <http://johndfenton.com/Rating-curves/Rating-curves.pdf>

Contents

1	Introduction	2
2	The complexity of simple hydraulic formulae	3
3	The power function $Q = C(h - h_0)^\mu$	4
4	Problems with logarithmic scales	9
5	Piecewise continuous interpolation	13
6	Approximation by global polynomials	17
7	Local approximation	28
8	Approximating splines	28
9	Results	33
10	Rating curve changes with time	41
11	The rating envelope – a generalisation	43
12	Rating shifts and the incorporation of later measurements	49
13	Conclusions	51
	References	51

1 Introduction

The problem of approximating stream rating data is important in water engineering, but despite the use of modern screen-based software, the underlying science seems often to be in a pre-computer state. In one approach the data is approximated visually, selecting points and varying a parameter so as to achieve two possibly conflicting requirements – to approximate the data by a curve and for that curve, unnecessarily, to be a straight line. Such a procedure is arbitrary and laborious, and in this era, surprising.

This paper takes positions which are contrary to current attitudes and practices. In several documents and on commercial websites one can find advocates of the use of open channel hydraulics in developing approximations to rating data. Whereas that field is an interesting and useful one that provides simple rough solutions to complicated problems and to understanding the processes at work, when it comes to formal approximation of data, it is of little assistance. We will see that in several places, errors have been made, not because of ignorance of hydraulics but because of the imposition of too much hydraulics and too little mathematics and data approximation. In this work the data is considered most important, and it is necessary to approximate that data with automatic methods that are accurate and simple to apply – reminiscent of the replacement of hand tools for spinning and weaving by mechanical ones.

It is surprising that there has been relatively little intellectual exchange between research organisations and professional practice in this area. Books and documents produced in universities, for example, repeat the standards of practice. There has been relatively little research in this area that has been published – in spite of the importance of the problem in river engineering. There seems to be little culture of questioning, rather there is an acceptance of the *status quo*. International Standards and Manuals seem

just to define and maintain the canon. Standards in particular often give formulae to estimate errors, without providing methods that are more robust.

2 The complexity of simple hydraulic formulae

Summary: *Flow formulae based on simple structures such as sharp-crested weirs and those based on uniform flow equations such as the Gauckler-Manning equation and Chézy-Weisbach equation have formed the backbone of open channel hydraulics. Here we consider flow formulae for two problems – a rectangular sharp-crested weir, and uniform flow in a trapezoidal channel to show that the formulae obtained for those idealised problems are more complicated than just a simple power function $Q \propto (h - h_0)^\mu$ and there is no technical reason to force rating curves to follow that over-simplified form.*

2.1 Weir formulae

Even if ideal conditions obtained (such as a small smooth rectangular channel terminated by a smooth planar end containing a single rectangular notch), the formula for discharge is a complicated function of geometry. For example, consider the standard formula for the flow over a sharp-crested weir:

$$Q = \frac{2}{3} C_D \sqrt{2gb} (h - h_0)^{3/2},$$

where C_D is coefficient of discharge, g is gravitational acceleration, b is length (transverse) of weir, h is stage and h_0 is the elevation of the sharp crest. The present author (Fenton 2015c) obtained the relationship, based on experimental results of Kindsvater & Carter (1957):

$$C_D = 0.589 - 0.008 \frac{(h - h_0)}{P} + \left(\frac{b}{B} \right)^2 \left(0.013 + 0.083 \frac{(h - h_0)}{P} \right),$$

where P is apron height (crest height above bed), b/B is relative fraction of stream width which the weir occupies. Over a realistic range of $(h - h_0)/P$ from 0 to 1 and b/B from 0 to 1 (full width), C_D varies between 0.58 and 0.68. Clearly as C_D is a function of $h - h_0$ we do not have a simple relationship $Q \propto (h - h_0)^{3/2}$, and the rating curve would not plot as a straight line on log-log axes. In all practical situations – it might be one of two or more as part of a larger structure and there might be piers between them, and so on – the geometric situation is more complicated than that from which such weir formulae were obtained. The knowledge of the real coefficient of discharge would be extremely approximate. And of course, other weir shapes such as V-notches and free overflows and broad-crested weirs form large additional families with different relationships between Q and $h - h_0$, and are probably not as well investigated.

2.2 Uniform channel flow formulae

If one has a situation where channel control prevails, and uses the Gauckler-Manning formula for example, the situation is more complicated. Backwater effects in open channels extend for a very long distance, so that downstream variation in cross-section, roughness, or slope affect the flow at any particular section. In any case, even if the flow were uniform, one usually has little idea of the cross-section, and even less of the slope or of the resistance coefficient, Manning's n . Even if one had an ideal situation where these were known, consider the Gauckler-Manning formula for discharge in a trapezoidal channel:

$$Q = \frac{\sqrt{S}}{n} W (h - h_0)^{5/3} \frac{(1 + \gamma(h - h_0)/W)^{5/3}}{(1 + 2\sqrt{1 + \gamma^2} (h - h_0)/W)^{2/3}},$$

where S is bed slope and stream slope in this uniform flow approximation, the horizontal bottom has a width W , the batter slopes are $\gamma: 1$ (H:V), and h_0 is the elevation of the channel bottom. Again, obviously the discharge does not follow a pure power function $Q \propto (h - h_0)^{5/3}$, and the rating curve would not plot as a straight line on log-log axes.

For overbank flow, the situation is more complicated, and there is even less reason to assume a simple power function with a straight line plot on log-log axes.

To conclude: simple hydraulic formulae provide a rough approximation to the stage-discharge relationship. However they generally cannot be trusted to be accurate to better than 5-10%, say, whereas the desired accuracy of a rating curve would be more like 1%. Possibly more important than any quantitative deduction here is the recognition that hydraulics shows that even in simple geometries, discharge does not vary simply like stage $(h - h_0)^\mu$, where μ is a constant. Accordingly there is no justification for invoking hydraulics to suggest that parts of the rating curve should be straight lines on a plot of $(\log Q, \log(h - h_0))$.

One might say that hydraulic formulae have been particularly misleading in causing people to believe in them unconditionally in determining the form of a rating curve. Worse would be to use hydraulic formulae without calibration, which would be just wishful thinking.

3 The power function $Q = C(h - h_0)^\mu$

Summary: The function has been elevated to a higher status than it deserves. We have just seen that it is an approximation anyway. Despite apparent simplicity, it is difficult to handle and there have been almost no computational methods for calculating it. Nevertheless, we discuss it here at some length and use some of its properties in later work. A computational method is developed. It is shown that the offset h_0 has little physical significance and the zero flow point is a myth. A simpler alternative for low flows is developed which approximates data well. Its significance is more in providing guidance for subsequent more-general methods.

3.1 The formula

The power function formula for discharge Q as a function of water surface elevation or stage h , arising from formulae such as in the previous section, is

$$Q = C(h - h_0)^\mu, \quad (1)$$

where C is a constant, h_0 is a constant elevation reference level, and μ is a constant with a typical value in the range 1.5 to 2.5. Obviously if $h = h_0$ the predicted flow is zero – and already a common error in the literature can be recounted, simple and obvious enough, but an error nevertheless, that when the term in the bracket is written $h + a$ then a is usually described as being the stage for zero flow. Of course it is not, that is given by $-a$. We also note terminological errors in the literature – this is not an "exponential" function, and neither do we like to think of it as a power "law". We are, unfortunately, being pedantic. Let us move on.

If we take logarithms of both sides of equation (1),

$$\log Q = \log C + \mu \log(h - h_0), \quad (2)$$

then on a figure plotting $\log Q$ against $\log(h - h_0)$, a straight line is obtained. A problem is that h_0 , the nominal zero flow point, is not initially known and has to be found. A much larger problem is, of course, that the equation is only a rough approximation, but because of its ability occasionally to describe almost

all of a rating curve roughly, it has acquired an almost-sacred status, and far too much attention has been devoted to it rather than addressing the problems of how to approximate rating data generally and accurately. The author lays the blame for the relative lack of sophistication in rating data approximation on that equation, where much modelling and computer software follow its procrustean dictates. It really has been believed to be a "law".

3.2 Numerical solution

No computational method is given in any Standard or textbook for the problem of fitting a function like equation (1) to a cloud of scattered data points. Commercial software seems to prefer laborious hand manipulation trial-and-error methods. The exception is the method of Lauffer (1996), which we shall briefly describe below.

In spite of our aversion to it, here we develop a couple of methods to fit the power function to scattered data. An automatic computational method which immediately springs to mind is to use standard modern optimisation software, such as the Solver module found in spreadsheets, to fit equation (1) or (2) to the data to determine the parameters C , h_0 and μ . Such software is generally capable of fitting a three or more parameter function to physical data, even if that function is nonlinear. However, that method does not always work so well for a function of the form of equations (1) and (2). The problem seems to be the highly embedded nature of h_0 and the fact that solution by iterative approximations deep in the software can give the problem of complex numbers if $h - h_0$ becomes negative anywhere in the course of iterative solution.

That problem is easily overcome if equation (1) is re-written as

$$h = h_0 + DQ^v, \quad (3)$$

where there are two unknowns occurring linearly, h_0 and $D (= C^{-1/\mu})$, and a third $v = 1/\mu$ whose effect is nonlinear but smooth. For a number of N data pairs (Q_i, h_i) for $i = 1 \dots N$, one substitutes Q_i into equation (3) to give an approximation to h_i and one calculates the error ε_i at each data point $\varepsilon_i = h_0 + DQ_i^v - h_i$ to give the total error the sum of the squares of the errors

$$e = \sum_{i=1}^N (h_0 + DQ_i^v - h_i)^2, \quad (4)$$

and then uses optimisation software to find h_0 , D , and v such that e is minimised. Of course, we can then obtain Q using

$$Q = \left(\frac{h - h_0}{D} \right)^{1/v}. \quad (5)$$

The author has found that solution by standard optimisation software, such as found in spreadsheets, works well. Such an EXCEL spreadsheet solution, using the SOLVER module, is given in <http://johndfenton/Rating-curves/Power-function-solution.xlsx>, which actually includes solutions in the form of both equations (1) and (3), in spite of the above warnings about the former.

Lauffer (1996) used an iteration procedure to fit equation (3), initially assuming $v = 1/2$, a typical value, and repeatedly solved for h_0 and D by linear least-squares and iteratively finding the solution which gave a best coefficient of correlation. The present author developed a simpler method using a standard least-squares procedure, eliminating h_0 and D from two of the three equations, leaving a third nonlinear equation for v to be solved numerically. However, the details of both these methods become complicated. It is simpler just to use the optimisation procedure described, being able to be set up quickly in a spreadsheet, for example.

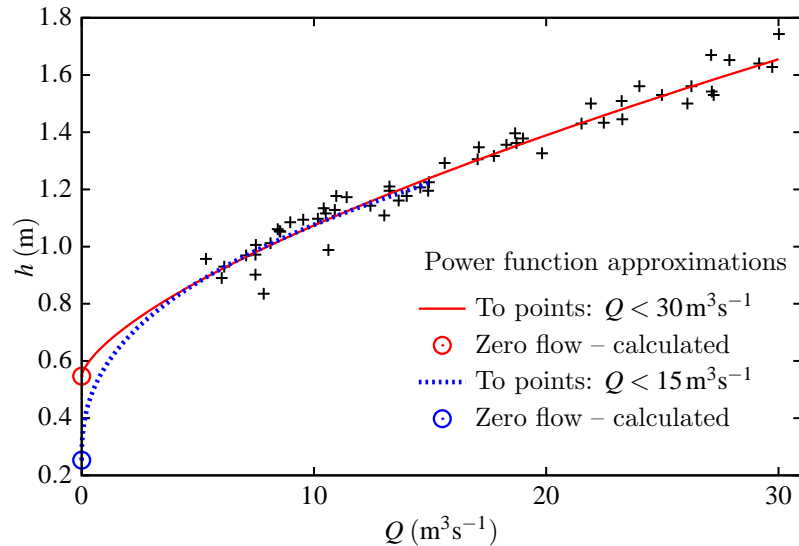


Figure 1: Two different approximations for low flows, giving effectively the same results within the data, but very different results for outside it and for h_0 . Data from USGS Station 02361500, Choctawhatchee River near Bellwood AL, USA, 2000-12-07 – 2015-05-22

3.3 The myth of the Zero Flow Point

There is a problem with interpretations of the results from fitting the power function. In particular this concerns the point of zero flow h_0 . In Figure 1 we show typical solutions using the above procedure for low-flow data points from the United States Geological Survey (USGS) Station 02361500 on the Choctawhatchee River near Bellwood Alabama, for 15 years of gaugings. We have converted all units from British to *Système Internationale* and used optimisation software to fit the power function. There are two curves shown, one fitted to points for $Q < 15 \text{ m}^3 \text{ s}^{-1}$, the other to all points for $Q < 30 \text{ m}^3 \text{ s}^{-1}$. It can be seen that in the regions of approximation the two curves apparently fit the data well (and practically coincide), but outside the data range the two curves diverge and both give quite different results for the zero flow point.

Clearly the values of parameters extracted are extremely sensitive to the data used, and in fact, there is almost an Uncertainty Principle in operation here: ideally, to calculate when the flow in a river is zero, one should use, not points for $Q \approx 30 \text{ m}^3 \text{ s}^{-1}$ as we have done here, but points for $Q \rightarrow 0$. In all cases we have seen, that is where results become highly irregular and one cannot reliably fit a function. The closer we get to actual zero flow, the less we are able to predict it accurately.

Our conclusion based on this and on other results we have seen is that the "zero flow point" is usually not a point at which flow in the river is zero, it is simply the value of h_0 to be used in the power function approximation, and we dismiss any consideration of extrapolation outside the data range. We cannot recommend using h_0 for any river calculations other than in the expression $h - h_0$ in the above approximations. Further, we assert that if zero flow has never been observed at a particular site, then zero flow is not a concept to be used there.

Or, as put succinctly by Lauffer (1996, top of p133, translated here from the German):

A measurable zero flow point does not usually occur. Also, even if this were the case, the published rating curve, like with all other points, will more or less not agree with it.

3.4 A simple alternative

We have already considered two different ways of writing the power function, equation (1): $Q = C(h - h_0)^\mu$ and re-arranged as equation (3): $h = h_0 + DQ^\nu$. We re-arrange this yet again and write

$$Q^\nu = a_0 + a_1 h, \quad (6)$$

where a_0 and a_1 are coefficients. In this form the right side could clearly be continued as a polynomial with higher degree terms in h . We will consider this at length in §6.

Chester (1986) advocated plotting rating curves with discharge raised to a power $\nu = 2/5$, based on the discharge formula for a triangular sharp-crested weir, suggesting that if the value of ν were correct, then according to equation (6) the points should fall on a straight line. The present author (Fenton 2001) compared discharge formulae over a family of sharp- and broad-crested weirs and uniform flow in channels, all varying from triangular through U-shapes to rectangular in cross-section, and concluded that plotting to the power $\nu = 1/2$ was a more representative average value. He used approximation to fit a 6th-degree polynomial to values of $Q^{1/2} = \sqrt{Q}$ for a single example and found good results.

Here we will examine the accuracy of writing simply $\sqrt{Q} = a_0 + a_1 h$ for low flows. The first reason is to develop a simpler method than the power function method above, but also, as preparation for approximation of greater generality using higher degree polynomials and an approximating spline method, using Q^ν .

To determine the coefficients a_0 and a_1 for a real data set, we can use linear least-squares. Here, we are using Q^ν as the dependent variable. Re-writing equation (4) we have

$$e = \sum_{i=1}^N (a_0 + a_1 h_i - Q_i^\nu)^2, \quad (7)$$

where now ν is assumed *a priori*. Differentiating e with respect to a_0 and then a_1 , setting both expressions to zero for an extremum and re-arranging gives the two equations

$$\begin{aligned} a_0 N + a_1 \sum_{i=1}^N h_i &= \sum_{i=1}^N Q_i^\nu, \\ a_0 \sum_{i=1}^N h_i + a_1 \sum_{i=1}^N h_i^2 &= \sum_{i=1}^N h_i Q_i^\nu, \end{aligned}$$

or, introducing the symbol S_{mn} for the general sum $S_{mn} = \sum_{i=1}^N h_i^m (Q_i^\nu)^n$, the equations are $a_0 N + a_1 S_{10} = S_{01}$ and $a_0 S_{10} + a_1 S_{20} = S_{11}$, with the explicit solution

$$a_0 = \frac{S_{20}S_{01} - S_{10}S_{11}}{NS_{20} - S_{10}^2} \quad \text{and} \quad a_1 = \frac{NS_{11} - S_{10}S_{01}}{NS_{20} - S_{10}^2}. \quad (8)$$

To test the methods we considered the low-flow ends of seven different data sets, each of which are described more fully in §9. We chose the upper boundary for low flows to be where the data seemed no longer to follow a simple smooth curve. Results are shown on Figure 2. In each case, we first tested the power function, equation (3), and used the optimisation procedure described in §3.2 to fit the data. On the figure the results are shown by blue dotted lines. The computed values of ν are also shown, which fall between values of 0.34 and 0.68. The lowest value, from the Ganges River, is unconvincing – the figure shows that there is already some consistent detailed structure in the variation. Next we fitted equation (6) with $\nu = 1/2$ using the explicit solution (8). The results are shown by solid red lines. This rather simpler approach seems to approximate the data just as well as the three-parameter power function. Hence we assert that there is nothing special about the power function or the deduced value of ν . We also tried a quadratic function, with an extra term $a_2 h^2$ in equation (6) and obtained an explicit solution like equation (8), albeit rather longer. There was little gain in accuracy.

However, if we wanted to apply this method to all the data points for a gauging station, it is to be expected that a polynomial of higher degree will be necessary, and that is what we will do further below.

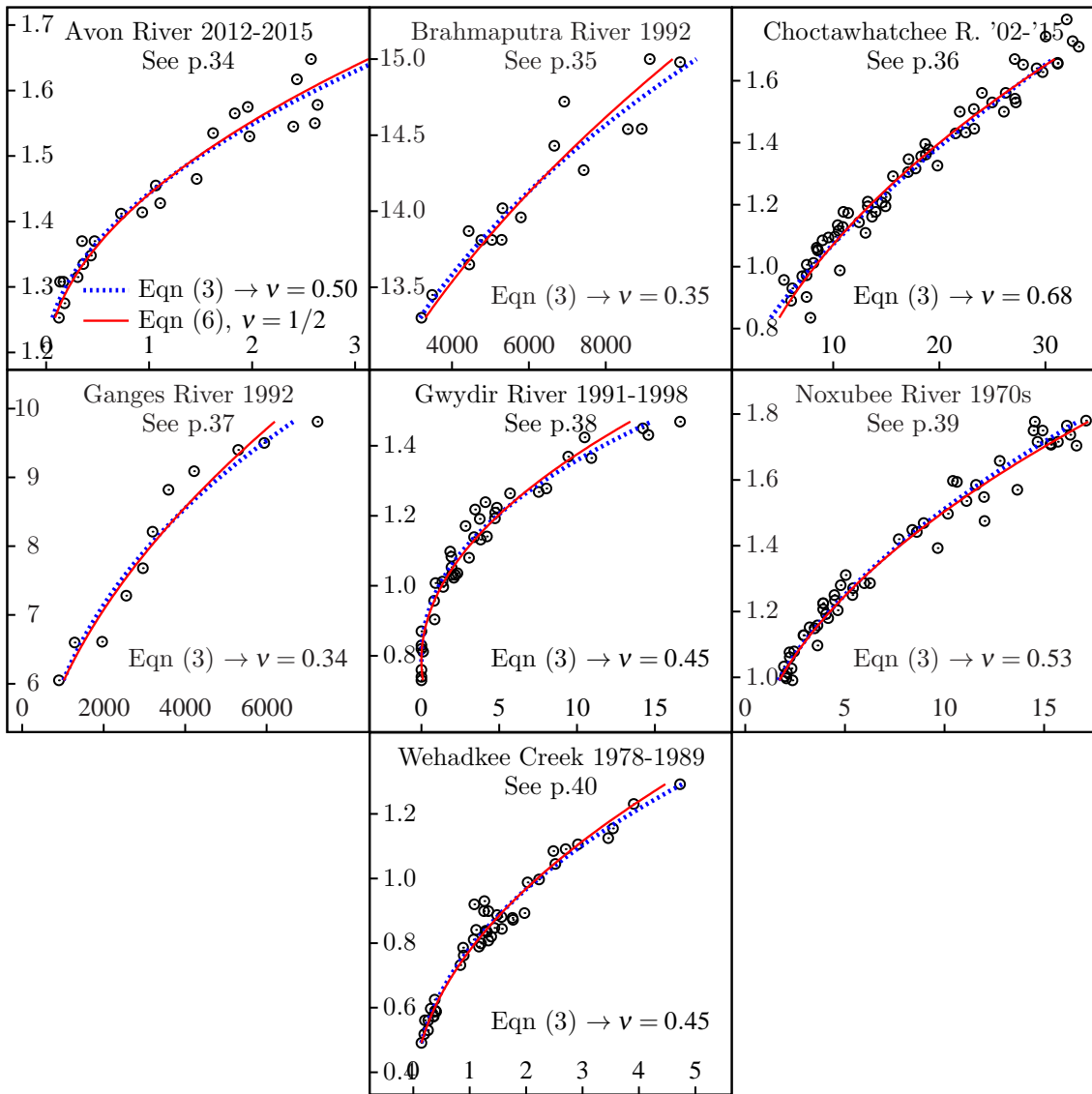


Figure 2: Two approximations to low-flow data, one using optimisation to fit equation (3), giving the calculated value of ν shown, the second using equation (6) with $\nu = 1/2$, fitting with the explicit solution (8). Horizontal axes are Q in m^3s^{-1} , vertical axes are h in m.

4 Problems with logarithmic scales

***Summary:** Some problems and errors associated with logarithmic scales, especially the stage scale, are identified. The largest of these problems is asserted to be due to confusion between the scales of the logarithm of the stage h and that of the offset stage $h - h_0$ and is at the core of current software. It is suggested that screen graphics programs should not be trying to solve rating approximation problems by varying h_0 interactively so as to approximate data and impose straight line variation over large parts of the rating curve. Instead, simple piecewise continuous linear approximation could be used, recognising that it is a data approximation problem and the power function does not necessarily apply. The possibly surprising result is obtained that on logarithmic axes of stage and discharge, $(\log Q, \log h)$ all straight lines have zero offsets and piecewise linear approximation of the data can be performed simply and automatically, almost without operator involvement. Finally, an elementary mathematical error in common software is noted in the labelling of screen displays of logarithmic stage axes.*

The approximation of rating data is made difficult by the fact that discharge Q might vary by a factor of 10^4 from the smallest to the largest and it is considered necessary to be able to approximate the data with a comparable relative error over the whole range. To represent the data on a plot, using the logarithm of discharge $\log Q$, is not unreasonable. However there is little physical reason also to use a logarithmic scale for h , as it typically varies at most by about 10m. The main reason for the traditional use of a logarithmic scale has been that if one actually plots $\log(h - h_0)$ against $\log Q$, if one chooses an appropriate value of h_0 , it is possible to obtain a straight line which is a satisfactory approximation to some of the data for small discharges and stages, or in some cases, for more of the data.

A problem is that the use of logarithms has caused problems of understanding, and complications, including one with many practical implications.

4.1 A minor problem

A small and relatively unimportant example is in Sauer (2002, p50), with much the same wording in WMO (2010, p.II.1-8) (reflecting the tendency in this field for material to be handed down from one source to another):

Logarithmic plots of rating curves must meet the requirement that the log cycles are square. That is, the linear measurement of a log cycle, both horizontally and vertically, must be equal. Otherwise, it is impossible to hydraulically analyze the resulting plot of the rating.

The author believes that to be simply not true. If one wanted to extract a gradient from a plot, one could extract all relevant information from say data pairs $(\log h_1, \log Q_1)$ and $(\log h_2, \log Q_2)$ one would take the actual values, one might measure a physical gradient on paper, but not, using contemporary technology, on a screen where any vertical distortion is possible.

4.2 Problems with the stage axis – and a simple solution

Sauer (2002, p50) writes:

A normal logarithmic scale (no offset) always should be used for the abscissa, or dependent variable. However, the ordinate scale should be adjusted, by default, by an amount equal to the offset defined for the primary rating being plotted. If multiple offsets are defined with this rating, and the user chooses to plot a continuous rating for the complete range of all segments, then the electronic processing system should default to the offset corresponding to the lowest segment of the rating to make the initial plot.

The word "always" in dictating the procedure for the discharge axis is a symptom of some things that have been wrong in the canonical procedures and publications in this field. For the stage axis, the "offset" that Sauer refers to is a value of stage h_0 such that if the power function $Q = C(h - h_0)^\mu$ is plotted as an approximation to the data on a figure with axes $\log Q$ and $\log(h - h_0)$, the result is a straight line, where the h_0 is determined by hand manipulations. This requirement, to approximate the data *and* to use a straight line of quite large extent seems often to be contradictory. Where the data is actually not well approximated by a straight line, thereby not following the power function, it leads operators into difficulties. It would have been better to have allowed them simply to approximate the data between the interpolation points with a curved line.

Screen-graphical software seems to allow different ranges of stage where different straight lines are used with different power functions, different offsets, and it seems that even the different ranges of data are plotted with those different offsets. Then when smaller lines are introduced to smooth the results, presumably they use the same local offset so that they are actually curved. It seems very complicated.

We suggest another way of using piecewise linear interpolation. Our aim here is to generate mathematical functions to approximate data. That data knows no "offset", and does not have to follow a power function. We do have to choose a sequence of interpolation points (Q_j, h_j) , $j = 1, 2, \dots$, so that locally each approximates the scattered points in its vicinity, for which we too would probably use clicking on a screen display of the data using whatever axes we thought reasonable.

The simplest reasonable approximation of the data over any part of it is a straight line. For us to recommend that sounds hypocritical in view of our previous criticisms of straight lines, but we use them purely as an approximating device, not believing we are following laws of hydraulics and accordingly not trying to impose straight lines where they are not appropriate. We would use as many as necessary, and so more where the data seems to be curved. We can compute the equations of those lines trivially. We can choose whatever axes we like for our piecewise linear interpolation, maybe one of Q , $\log Q$, or \sqrt{Q} (Fenton 2001), and one of h or $\log h$. For the moment we choose $\log Q$ and $\log h$, so that the linear interpolating/approximating function $Q(h)$ in $(\log Q, \log h)$ space between any two points $(\log Q_j, \log h_j)$ and $(\log Q_{j+1}, \log h_{j+1})$ is

$$\log Q = \log Q_j + \frac{\log h - \log h_j}{\log h_{j+1} - \log h_j} (\log Q_{j+1} - \log Q_j). \quad (9)$$

That would be enough for our purposes, so that we could plot the rating curve and use it to calculate flows from routine stage measurements. However, out of interest we manipulate that equation to give an equation for Q . We find it is just the simple power function

$$Q = Ch^\mu, \quad (10)$$

where

$$C = \frac{Q_j}{h_j^\mu}, \quad \text{and} \quad \mu = \frac{\log(Q_{j+1}/Q_j)}{\log(h_{j+1}/h_j)}. \quad (11)$$

Thus the local interpolating function, linear on logarithmic axes, equation (9), is simply the power function equation (1) – but with zero offset $h_0 = 0$! We have the obvious but possibly surprising conclusion that $h_0 = 0$ for *all* straight lines on $(\log Q, \log h)$ axes.

We believe that the laborious fitting of straight lines by adjusting the offset h_0 on $(\log Q, \log(h - h_0))$ axes has been a mistake by the industry. The alternative, computing a series of piecewise continuous straight lines just to interpolate a data set is actually easy – one just has to set all the interpolation points $(\log Q_j, \log h_j)$ for $j = 1, 2, \dots$ and use equations (10) and (11) between each pair, without any adjustment for offset. Even if piecewise linear is a relatively coarse and a not so aesthetic way of going about the problem, it is probably quite justified in view of the scattered nature of the data (and has been widely used in the industry in the form of straight lines on the $(\log Q, \log(h - h_0))$ axes).

We do need to examine how our piecewise linear interpolation performs. In books and standards, the examples chosen are such that in some places for parts of the rating curve, especially at low flows,

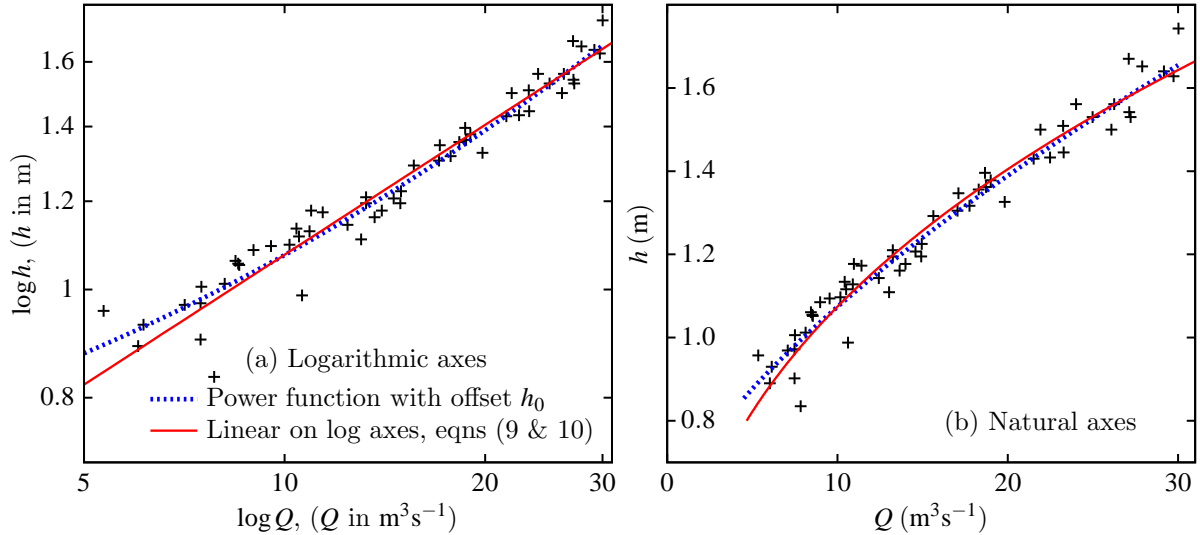


Figure 3: Comparison between linear function on logarithmic axes $h_0 = 0$ and the power law with h_0 fitted. Data: low flow results for USGS Station 02361500, Choctawhatchee River near Bellwood AL, USA, 2000-12-07 – 2015-05-22

the power function with a finite value of h_0 is expected give a more concise representation, with the hidden suggestion that one is following laws of hydraulics. Here we test the method with two problems. The first is for low flow data, where the power function with finite h_0 is widely believed to be useful. We calculated the power function approximation using the least-squares method we described in §3.2. Instead of calculating the "linear" approximation (10) by just using two interpolation points, we actually calculated it using a linear least-squares approximation to remove our bias. Results are shown on Figure 3, part (a) using the logarithmic axes where the straight line approximation is clear – and is just as good as the power function. Part (b) on natural axes, shows the expected curved nature of the data and of course the simpler function is also just as good.

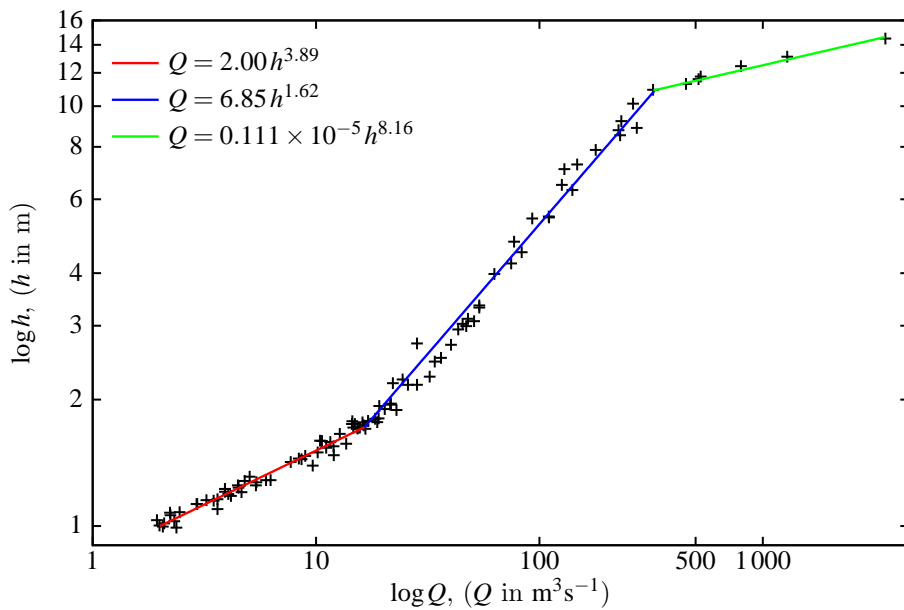


Figure 4: Piecewise linear interpolation on logarithmic axes for the Noxubee River near Geiger, AL, USA, USGS Station 02448500, 1970s

As further evidence for the applicability of the method, in Figure 4 we present the rating data and three simple linear approximations for the Noxubee River near Geiger, AL, USA, USGS Station 02448500,

for all gaugings from the 1970s. This is a remarkable data set, as for the three separate types of control, each is quite linear on the log-log axes – actually not something that we generally expect to this extent.

We believe these examples justify our hopes that straight lines on $(\log Q, \log h)$ axes are good enough for approximation and we do not have to use hand-manipulation methods. In §5 we will outline what we have done here in a more general context of *piecewise continuous interpolation*, as a simple way of generating rating curves.

4.3 Incorrect enumeration of logarithmic axes

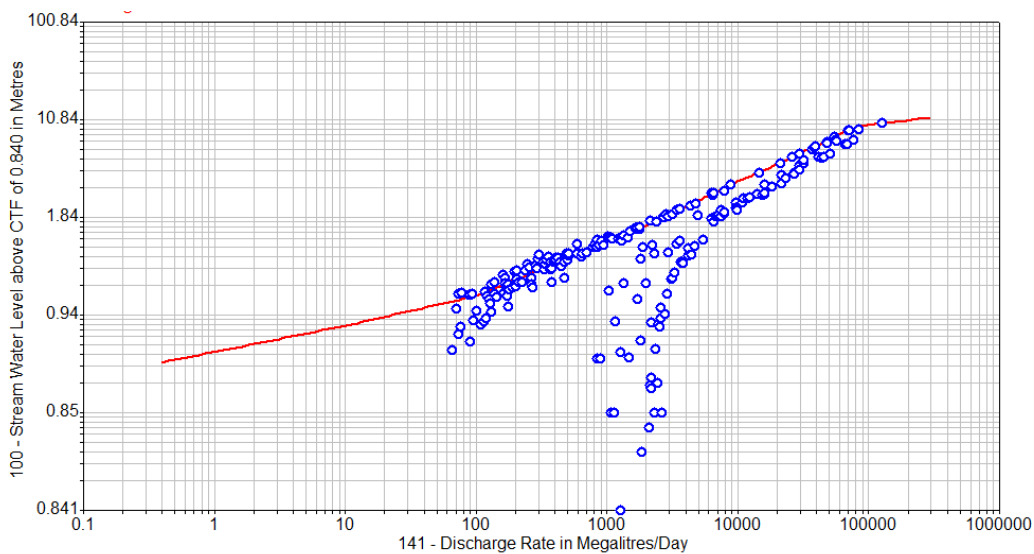


Figure 5: Screen-shot of commercial software, showing rating data and an ambitious rating curve, May 2015

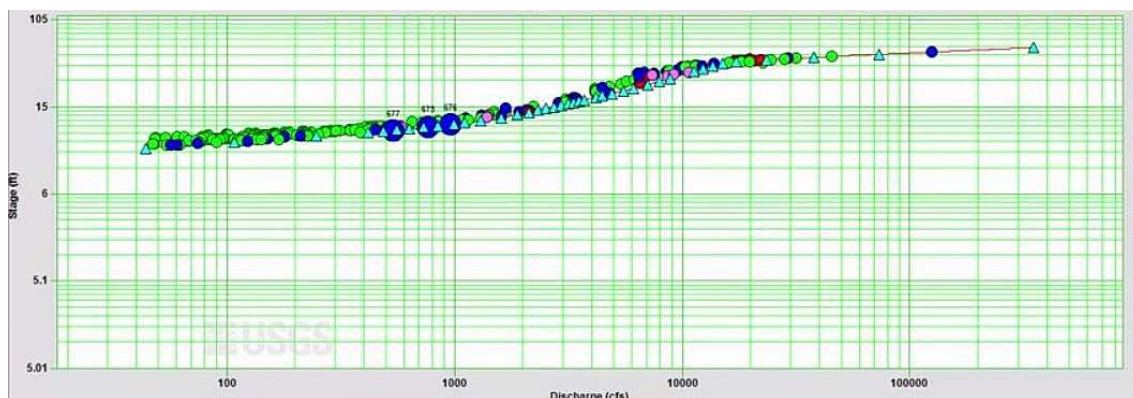


Figure 6: Screen-shot from training video - fitting straight lines to data segments

Now we consider two examples where confusion about the nature of logarithms may have led to significant errors in the calculation, presentation and use of rating curves. Again, this error seems to have come about because of confusion about offsets, leading to the mathematical nonsense of the addition of a constant to decadal markings on logarithmic stage scales.

Figure 5 shows rating data and a rating curve from a website of an Australian water agency apparently using well-known software. One problem is that the rating curve has been extrapolated far too much beyond the low flow end of the data. Another slight problem is that the units of discharge are the quaint non-*SI* Australian water industry ones of Megalitre/day ($86.4 \text{ ML/d} = 1 \text{ m}^3 \text{ s}^{-1}$). The real problem, however, is one that the current author has encountered on the sites of two other Australian state water

agencies, using the same software. Here the decadal markings on the vertical scale are 0.841, 0.85, 0.94, 1.84, 10.84, and 100.84, given by the sequence $10^{n-3} + 0.84$, where n is the number of the decade, so that we go from $n = 0$ at the bottom of the graph, with $10^{0-3} + 0.84 = 0.841$ to $n = 5$ at the top of the graph with $10^{5-3} + 0.84 = 100.84$. A similar problem is shown in Figure 6, a screenshot from a training video from an important water agency in the USA in units of feet and cubic feet per second. One can make out the decadal markings on the vertical axis 5.01, 5.1, 6, 15, 105, which correspond to the sequence $10^{n-2} + 5$ for n from 0 to 4.

In each case, the addition of a number, 0.84 and 5, is incorrect. For a logarithmic scale to have significance, distance along the stage axis, proportional to $\log h$, must increase by the same amount in each decade, that is for two decadal markings, h_{n+1} and h_n , the quantity $\log h_{n+1} - \log h_n = \log(h_{n+1}/h_n)$ is constant for all n , such that the ratio of any two corresponding heights is a constant. The sequences shown in the figures, with additive constants, do not satisfy such a relationship, and accordingly the decadal markings, and any deductions from both logarithmic scales, are wrong.

It may be that the numbers on the decadal markings in both cases shown are merely labels that have been calculated for convenience (although still not very convenient – interpolating a value of, say 1.23 m between decadal markings of 0.94 and 1.84 on such a non-uniform scale would be difficult). One wonders what the implications of this are for the calculated and tabulated rating curves of water agencies around the world that use software with such an elementary error, and whether it is made more deeply internally than just labels on a displayed axis.

5 Piecewise continuous interpolation

Summary: It is shown how piecewise linear interpolation can be simply and automatically implemented to generate rating curves. Such linear functions are quite satisfactory, however if greater smoothness of the rating curve is desired, then spline interpolation using cubic polynomials can be used, and is shown to give very good results.

Consider now the general problem of rating curve generation, where we have a number of scattered data points and we want to obtain a function which represents those points. There are two main approaches that we can take. One is *approximation*, obtaining a function that passes through the assembly of the data points, (Q_i, h_i) , $i = 1, 2, \dots$, approximating them in a least-squares sense such that the parameters of that function, for example the coefficients in a polynomial, minimise the sum of the squares of the differences between the function and the data points. This is capable of automation and further below we will spend a lot of effort describing and developing robust methods to do that. The second approach is *interpolation*, finding a function that passes through a usually smaller number of data points that follow a smooth variation. To apply interpolation in rating curve representation, one needs to insert data points for that process so that the result locally represents the scattered data points. This would be done, for example, by plotting or screen-graphical methods. For all the methods in this section, we assume that this has been done, giving another set of points, interpolation points, for which we use the same symbols h and Q as for the actual data points, but will use subscripts $j, j + 1$, etc.

5.1 Piecewise two-point interpolation

Here it is assumed that we will be performing piecewise interpolation, where a different function is used between each two inserted interpolation points, the function on either side agrees in value at the common point but there is no continuity of derivatives. This is very simple, and we have already applied it in one form in §4.2.

5.1.1 Linear interpolation in natural (Q, h) space

This is simply a straight line on an (Q, h) plot, passing between two points (Q_j, h_j) and (Q_{j+1}, h_{j+1}) :

$$Q = Q_j + \frac{h - h_j}{h_{j+1} - h_j} (Q_{j+1} - Q_j). \quad (12)$$

There is actually nothing wrong with this most simple of all approaches, as there is nothing special about the use of straight lines on logarithmic axes, other than that they can be used for larger intervals.

5.1.2 Linear interpolation in (Q^v, h) space

Here instead of Q , a scale of Q^v is used, where v is an exponent smaller than 1, as described in §3.4. We consider it just in the context of piecewise linear interpolation between two data points such that we would write

$$Q^v = Q_j^v + \frac{h - h_j}{h_{j+1} - h_j} (Q_{j+1}^v - Q_j^v). \quad (13)$$

At the low-flow end we have seen this work well for $v = 1/2$. However as a general method it does not work particularly well, with, in addition to the expected gradient discontinuities, also curvature discontinuities, which make it appear worse.

5.1.3 Linear interpolation in $(\log Q, \log h)$ space

We have presented and discussed this and its ramifications earlier, in §4.2, equations (9) and (10). It seems that this would be a good and simple general method of generating rating curves. It has gradient discontinuities at interpolation points, so one might say that it is a relatively coarse and a not so aesthetic way of going about the problem, however it is probably quite justified in view of the scattered nature of the data. In fact, it is the standard already used in the industry, but implemented using offsets.

We now turn away from piecewise linear interpolation and consider a smoother method which uses polynomials of higher degree and imposes rather more continuity at interpolation points.

5.2 Spline interpolation

We consider a well-known general method for the interpolation of an arbitrary series of data points, using splines, that are polynomials of low degree which span successive parts of the data. We will find that they give very good results, provided a certain procedure is followed at the ends of the data. What we will be doing here still requires the placement by an operator of a number of computational points, such that each is in the middle of the local data points. The method seems to be known in rating curve trade literature, but one might have thought it would be more widely used.

Splines are a sequence of polynomials of low-degree, such as second or third, but with a higher degree of continuity of function and derivatives at data points than we considered in the previous section and which did not work well. The method is described in many books (for example, Conte and de Boor 1980, and de Boor 2001) and included in software packages. The physical interpretation and the name of cubic splines is familiar to civil engineers, for it comes from a draughtsperson's flexible strip or "spline" which can be used to fair smooth curves between points. If the strip is held in position at various points by pins, then between any two of those pins there are no lateral forces acting so that the shear force in the strip is constant, the bending moment varies at most linearly, and hence by beam theory (for sufficiently small deflections) the strip takes on a cubic variation between the two points. As the variation of moment is different between other pairs of points, other cubics will apply there. However, because the shear force and bending moment are continuous across each pin, then the first and second derivatives are continuous

across the pins, or interpolation points. With four unknown coefficients for the cubic in between each pair of points, and the requirement that each of the two cubics, to left and right of each interpolation point, must interpolate at that point and must have the same first and second derivatives, almost enough equations are obtained for all the four coefficients of each cubic.

It is necessary, however, to specify two more conditions. This may be by specifying the first or second derivatives at the two end points, as in Conte and de Boor (1980, Section 6.7). In general, there is no particular reason at all why the first or the second derivative of the interpolating spline should take on a particular value at the ends, and in almost all presentations and software, with the exception of de Boor (2001), the method suffers from this disadvantage and is not as accurate at the ends as it might be.

A better way to obtain the two extra conditions is to use the "not-a-knot" condition at the first and last interior points (de Boor 2001), where it is required, in addition to the first and second derivatives agreeing to left and right of the interpolation point, that also the third derivatives agree. The physical significance of this is simply that a single cubic interpolates over the first two intervals and another over the last two intervals. No arbitrary assumptions have been introduced, and it can be shown that the method is more accurate than those that require end specifications of derivatives.

Now we describe the application of spline interpolation to three sets of rating data. For each we allocated knot points – this much still requiring human interaction – and performed the spline interpolation. We did two interpolations for each case, one using the actual discharges Q_j , and the other, using $Q_j^{1/2} = \sqrt{Q_j}$ which has the effect of making the data form an almost-straight line for small discharge and for the highest flows to make numerical values smaller, so that the splines do not have to interpolate such large values.

Results are shown in Figure 7. Part (a) is for Pallamallawa on the Gwydir River, NSW, Australia, Site number 418001, 1991 – 1998. We had thought that this might be a difficult problem, as there is a large gap in the data, but it can be seen that the splines, with only 5 interpolation points, approximate the data well. Interestingly, for the lowest flows, which were really very small (zero, actually!), interpolating the actual discharges Q_j gave the irregularity shown. This could have easily been overcome by using one or more interpolation points at the low flow end, but we wanted to show the phenomenon here. The results show that using $\sqrt{Q_j}$ gave better results, which we have often found throughout this work, whether for interpolation or approximation.

Figure 7(b) is for the same data we used in Figure 4, where we simply approximated the data with three straight lines. The spline approximation, admittedly with 9 interpolation points, is more accurate. Both Q_j and $\sqrt{Q_j}$ methods worked well.

Figure 7(c) is for the Ganges River in 1992, with data taken from Mirza (2003). The data is relatively complicated with an identifiable structure. We chose to approximate that structure with 11 points, rather than to smooth it with fewer, and the spline method worked well.

We conclude that the approximation of rating data by splines works very well indeed. One has only one goal, to represent the data as well as possible, and considerable flexibility can be followed in the placement of interpolation points. Now, however, for the rest of this report we will concern ourselves with approximation methods, which can be programmed to function almost automatically, with a minimum of human judgement and activity required.

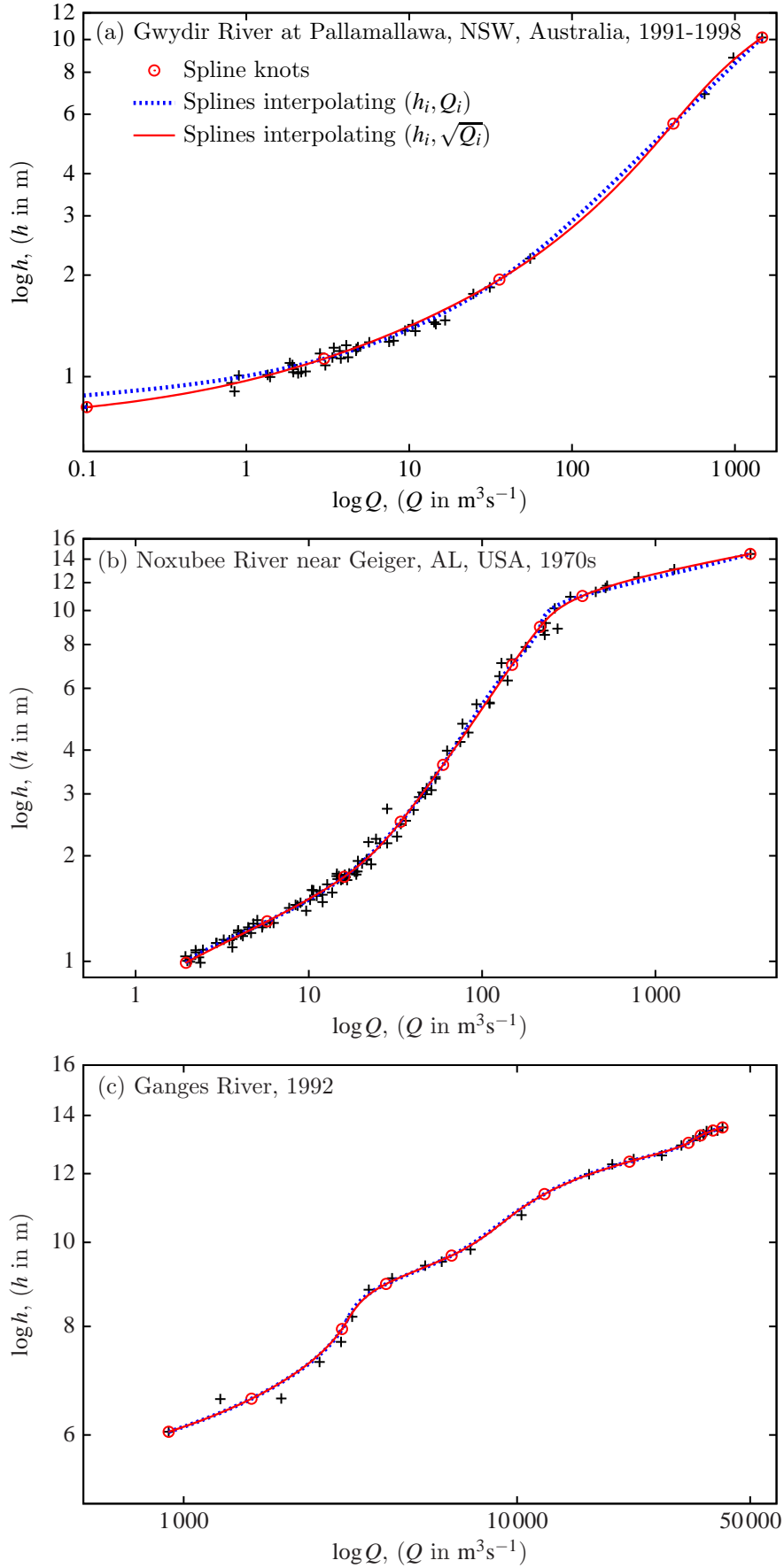


Figure 7: Three examples of rating curve generation by cubic spline interpolation between the knot points shown, determined by the author

6 Approximation by global polynomials

Summary: *Interpolation, which was just described, requires the insertion of a number of points through which the rating curve must pass, relying on an operator’s judgement where to place the points. Approximation by least-squares is a more powerful and automatic way of representing rating data. There are potentially several problems with such a method as it is presented in books and standards. Those problems may well be the reason that it seems not to have found favour. They are described and a theory and method presented which overcomes them. The only arbitrariness requiring operator involvement is in the degree of polynomial to be used. Good results are obtained.*

Here we abandon piecewise-continuous methods and calculate a curve given by the same function (“global”) over all water levels. The expression for Q as a polynomial in h is written, following several standard sources (for example, International Standard 7066-2 1988, Herschy 2009):

$$Q = a_0 + a_1h + a_2h^2 + \dots + a_Mh^M = \sum_{m=0}^M a_mh^m, \quad (14)$$

where the degree M of the polynomial might be something like 4 – 10, and the a_m are coefficients to be found by approximation procedures such as least squares.

International Standard 7066-2 (1988) is the most scholarly presentation of global polynomial approximation in the context of rating curves. It describes in principle how to do the least-squares analysis, but only for the trivial case of a quadratic, $M = 2$, whereas the general case can be presented concisely. It then mentions computational methods for solving the *Normal Equations* which result from the least squares analysis, and mentions that there may be computational problems, but does not explain what they are. The reason is, in fact, well known to numerical analysts – the Normal Equations are particularly prone to ill-conditioning and numerical inaccuracy, because they tend to be almost multiples of each other so that solutions are poorly defined. Next, however, the Standard remedies this by presenting a scheme where more robust equations are obtained by *Orthogonal polynomial* curve fitting, and presents a computer program. Below, however, we will show how that is not a solution to one of the problems of the method.

Reading books and mentions on water industry websites and reading between the lines it seems that the approximation by global polynomials, despite its promise, has never really been successful, and is usually only implemented to something like $M = 5$. Herschy wrote in the first edition of his book in 1985, and 24 years later again in the third edition Herschy (2009, p195): “However some user experience is still required with this method before it is accepted as an alternative to the existing methods”. That suggests a lack of faith in it.

There are, in fact, some simple but powerful reasons for any problems associated with the method, which we will now describe.

6.1 Reasons for problems with global approximation, and some solutions

6.1.1 Insufficient diversity of the monomial functions

Equation (14) shows how the polynomial suggested in books and standards is made up of a linear combination of the monomial basis functions h^m , each weighted with a coefficient a_m . For any such approximation to be robust and accurate, it is highly desirable for the individual basis functions to have dissimilar behaviour from each other so that irregular variation can be described efficiently. In this respect, the monomial functions are not very good, as they all look rather like each other for large h and for $m = 2$ or greater, and with no distinct regions of higher curvature.

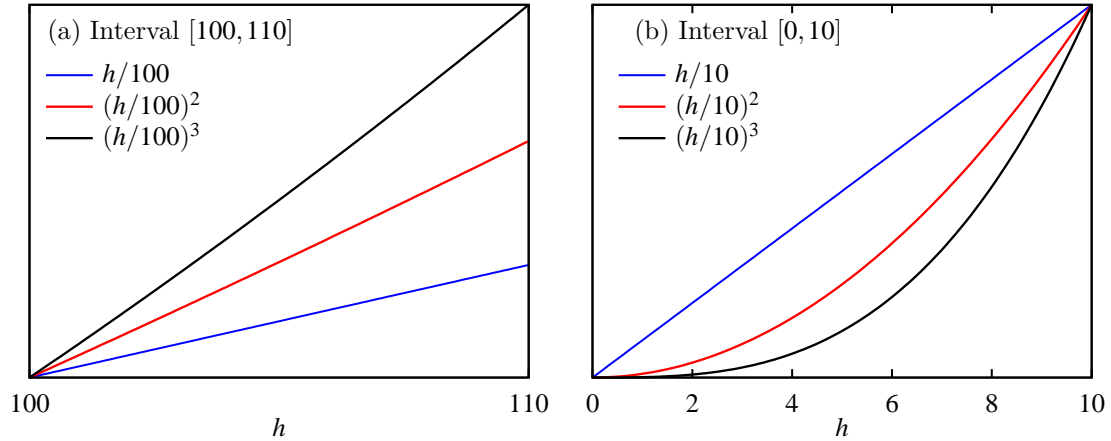


Figure 8: Comparison of the first few monomials on the intervals [100, 110] and [0, 10]

Range of stages small compared with actual values: this is usually not the case, but it must be warned against. The problem occurs, for example, if the stage datum is sea level, so that the actual stage values at the station might be something like 100m, with a probable maximum range of 10m. Figure 8(a) shows that situation, with three low degree monomials plotted over the range [100m, 110m]. They all look like each other, little different from straight lines. This would make a linear combination of them approximating any data with curvature difficult, and the resultant polynomial would have large coefficients oscillating in sign. Of course, almost everywhere in practice a local datum is used, but one can never be sure, and this shows how that should be avoided. It is strongly recommended that, at least, $(h - h_{\min})^m$ should be used – subtracting the minimum of the data points, h_{\min} , and using that in the basis functions. Figure 8(b) shows, effectively what is usually done, where the lowest points are close to the datum and the range of heights goes from about 0 to as much as 10m. The figure shows that these first few monomials at least show greater diversity of behaviour than in the interval [100m, 110m], with a better ability to represent data or functions.

Large numerical values of the independent variable h : the author (Fenton 1994) has described this in the case of civil engineering problems of interpolation, but for the present approximation problems the principle is the same. The concern there was for roads, railways and rivers, where the independent variable (probably a linear distance x there, but h here) can be very large. Even for river heights specified relative to a local datum, where the range might be typically only 0m to 10m, the contribution of the last term in the polynomial in equation (14) at the highest stage would be $(h_{\max} - h_{\min})^M$, or for $h_{\max} - h_{\min} = 10$ m and an 8th degree polynomial $M = 8$, a value of 10^8 so that the corresponding coefficient a_8 would have to be very small and specified accurately. Internal to a computer program this may not matter, but if numbers are transferred between software or people, much care would have to be taken. If the units were specified in feet, such as in Liberia, Myanmar, or the USA, with larger numerical values, it might be more of a problem. Even more potentially fragile would be the example of a rating curve where the stage is expressed in centimetres, which might be a common practice because the values are integers. This is used in Viet Nam, and in Morgenschweis (2010, p384), where the numerical range was 0cm to 700cm, possibly causing that author to note, in the context of global polynomial approximation, the restriction that "usually a polynomial of 3rd or 4th degree is enough".

It is worth mentioning here that this problem of poor conditioning of a polynomial formulation can be almost accidentally invoked. There is a facility in the spreadsheet EXCEL by which "trendlines", which are actually approximations, can be simply added to sets of data points. Internally the program is very robust, and the plotted curves are accurate. Often one needs the actual function which EXCEL has obtained, and that can be displayed on the figure as well. However, that facility has a flaw, in that it displays the formula in expanded form such as equation (14) with the problems that entails if the

independent variable is numerically large, so that round-off errors might be large. In the program one can change the number of digits shown, but the problem remains.

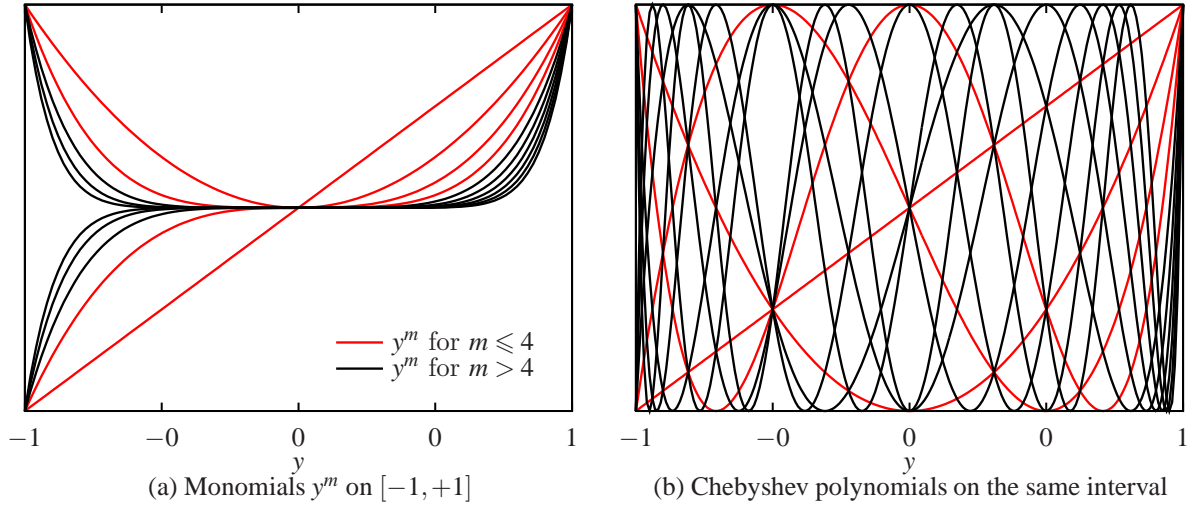


Figure 9: Comparison of the first 10 monomials and Chebyshev polynomials on $[-1, 1]$

Possible solutions – which we will find not to be solutions The problem of large numerical values could be overcome by re-scaling the interval of approximation from $[h_{\min}, h_{\max}]$ by using a variable $y_* = (h - h_{\min}) / (h_{\max} - h_{\min})$ to bring it to $[0, 1]$. Over this interval the first few monomials $1, y_*, y_*^2, y_*^3, \dots$ all have an order of magnitude of 1 and appear different from each other and so have better powers of approximation, as suggested in Figure 8(b). However for higher degree polynomials Figure 9(a) provides an explanation for the fragility even of these y_*^m , when m is large. If we just consider that part of the figure for which y is positive, namely y_* here, it can be seen that for $m > 4$ the monomials tend to differ little from each other, which again would require large coefficients to describe arbitrary variation. We found this within our computer program that this scaling had no advantages, which we will show further below.

Scaling to the interval $[-1, +1]$ using $y = -1 + 2(h - h_{\min}) / (h_{\max} - h_{\min})$ and using y^m as the basis functions should be better. Considering also the values of y for which y is negative on Figure 9(a), it can be seen that at least for m odd and y negative the monomials become negative, so there is a little more diversity of behaviour and these basis functions might be a little better than just using y_* . As we will see, they are indeed a little better, but that is all.

The best solution – Chebyshev polynomials: The next best approach, and in fact the *non plus ultra* of global numerical approximation, is the use of orthogonal polynomials that all differ from each other in behaviour. They have the property that under a process of weighted integration or summation over a special set of point values, the product of two such polynomials, one of type m , the other of type n , say, is zero for all $n \neq m$, only giving a contribution for $n = m$, and leading to the name "orthogonal". In our case we sum over all the scaled stages y_i of a data set, so that the sums $n \neq m$ over all the points are not zero but they are relatively small and the behaviour is quasi-orthogonal at least. This makes their behaviour much better than using other functions.

Chebyshev polynomials $T_m(y)$ use the variable y defined in the previous item, in the interval $[-1, +1]$. They can be simply evaluated by $T_m(y) = \cos(m \arccos y)$, or one can evaluate recursively from $T_0(y) = 1$, $T_1(y) = y$, and for all $m \geq 2$, $T_m(y) = 2yT_{m-1}(y) - T_{m-2}(y)$. Chebyshev polynomials are formally the most robust approach, and this is shown dramatically in Figure 9(b), where the diversity of behaviour suggests a better ability to represent data or functions. We will find that this is borne out by our computational results.

6.1.2 Regions of rapid variation, Runge’s phenomenon, and end oscillations

Even using Chebyshev polynomials is still not enough to overcome the next reason why global polynomial may not work very well. If the data to be interpolated or approximated has a region of rapid variation then because the global function has to approximate both that region and elsewhere where variation is slower, the interpolation or approximation can be poor globally. This is known as *Runge’s phenomenon*. The consequences can be serious and surprising, such that increasing the degree of approximation can simply make the problem worse. It is well-known that oscillations are more likely at the ends of the interval.

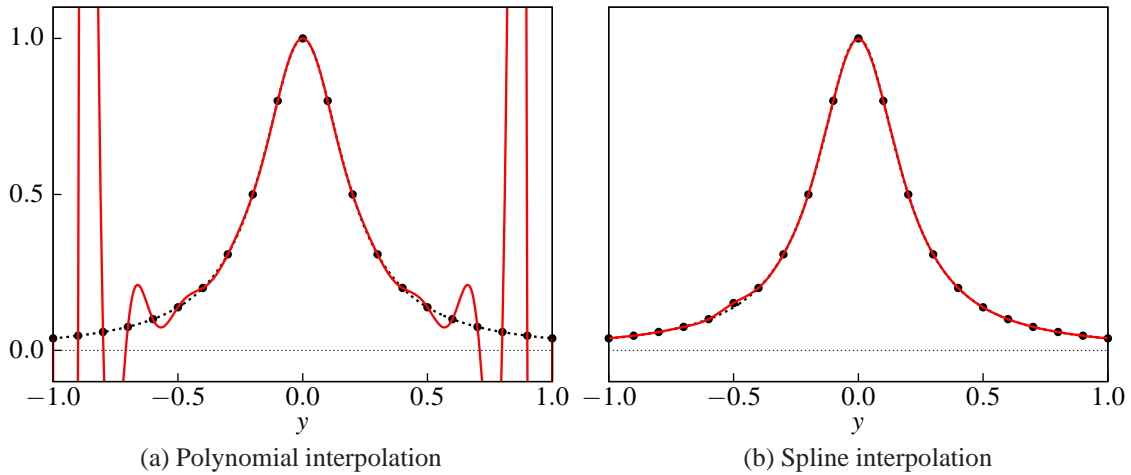


Figure 10: Interpolation of Runge’s function $1/(1+25x^2)$ with a sharp crest using 21 data points and a 20th degree polynomial. For the spline case, a data point at $x = -0.5$ was raised slightly – the effects are quite localised.

Figure 10(a) shows this dramatically for the global polynomial interpolation of a function $1/(1+25x^2)$ (on our recommended interval of $[-1, +1]$) devised by Runge to show that the region of rapid variation near the crest has destroyed the accuracy in the slowly-varying region away from the crest. The global nature of the approximation, giving good agreement near the crest, has meant that elsewhere the agreement is terrible, and taking more terms only makes results worse.

A series of Chebyshev polynomials would be no solution to the problem, because they would give exactly the same results, it is just that they would be able to be computed more robustly. This is why Orthogonal polynomial curve fitting described in International Standard 7066-2 (1988) would be of not much assistance. In rating curve approximation, regions of rapid variation (high curvature) can occur for the lowest flows dominated by a local control, or in cases such as a transition between local and channel control, or indeed anywhere.

Examining the Chebyshev polynomials drawn in Figure 9(b) it can be seen that in the limit as $y \rightarrow +1$, and to a lesser extent $y \rightarrow -1$, even they start to resemble each other, and it is no surprise that approximation at the ends of the interval of approximation is often less good, with oscillations, as we have seen in Figure 10(a).

Figure 10(b) shows the use of piecewise-continuous interpolating splines, overcoming all of the problems of global approximation we have mentioned. In §8 we will present a method for rating curve approximation based on splines.

6.1.3 Homoscedasticity and scaling of the dependent variable

In statistics papers concerned with approximation, more technically known as regression, there is much concern with homoscedasticity, or the requirement that the local variance of the points about the fitted curve be constant over the whole range. Otherwise, the Gauss-Markov theorem shows that least squares as we use it will result in bias of the results. In our case of rating data approximation, we have to deal with problems where at the lower end with Q of the order of $1 \text{ m}^3 \text{ s}^{-1}$ there might be scatter of at least 10%, so, $\pm 0.1 \text{ m}^3 \text{ s}^{-1}$, while for very large flows $1000 - 10000 \text{ m}^3 \text{ s}^{-1}$ there are likely to be few points and considerable scatter, of say, $\pm 1000 \text{ m}^3 \text{ s}^{-1}$. Presumably local variance changes by a similar ratio. We cannot hope to satisfy homoscedasticity. While there are sophisticated data transformation techniques available, we are going to use a couple of simple transformations, obvious from the hydraulics of the problem, which reduce the relative variation of the dependent variable and we find that we can obtain global approximations that can describe flows of $1 \text{ m}^3 \text{ s}^{-1}$ at one end and $10^4 \text{ m}^3 \text{ s}^{-1}$ at the other end. We will find that by simply using least-squares we can obtain good plausible results for all problems we have considered.

When this document was almost ready we met a professional statistician at a conference. She, (Ilaria Prosdocimi, Personal Communication, 2015) noted that the variance only really matters when you want to compute the significance of the coefficients or the confidence/prediction error of the estimates, but not so much the estimated values, which is what we are concerned with.

The first transformation we consider is to use $\log Q$ as the dependent variable and to use as basis functions, monomials $(\log(h - h_0))^m$ where h_0 is the zero flow point as calculated in §3.2. This obviously brings the range of dependent variable from a range of $10^0 - 10^4$ to something like $0 - 4$. We will see that this is not as good a solution as one might think, as the expansion of the region at the low flow end means that too much computational effort is spent in approximating that, at the cost of approximating the now more rapid variation in the high flow regions.

The second transformation we use is that of Q^v , where v is an arbitrary exponent ($v < 1$), that one could choose *a priori*, as described in §3.4. We will examine use of that transform here, in fact only using $v = 1/2$, and it will be shown consistently to give better results than fitting to Q .

Prosdocimi also observed that such transformations are well known, pointing out the problem from a statistician's point of view that the expected value of the transformed quantity is not the same as the transform of the expected value.

6.2 Using global approximation to generate rating curves

6.2.1 Least-squares theory

Despite our description of the potential fragility of global polynomial approximation, we now present it, in view of the above discussion, along with a simple method for implementation, to obtain rating curves using this global approximation in a least-squares sense. We are going to consider different forms for the nature of the approximation, so we will re-write equation (14) in a generalised form:

$$F(Q) = \sum_{m=0}^M a_m f_m(h), \quad (15)$$

where $F(Q)$ is a function of Q and the $f_m(h)$, functions of h , are the *basis functions*, linearly combined as shown. The simplest but not the best approach is that of equation (14) with $F(Q) = Q$ and $f_m(h) = h^m$.

We calculate the total sum of the squares of the errors e of the approximation over N data points (h_i, Q_i) for $i = 1 \dots N$ from equation (15):

$$e = \sum_{i=1}^N w_i \left(\sum_{m=0}^M a_m f_m(h_i) - F(Q_i) \right)^2, \quad (16)$$

where we have introduced something potentially useful, the weight w_i for each rating point, giving us the freedom to weight some points more if we wanted the rating curve to approximate them more closely, or others to be set to zero such as if we wanted to eliminate points obtained from different time periods to examine the effects of shifting control. Often, however, all the w_i will be 1. There is no need for data files to be sorted in increasing h_i or Q_i , which is useful if they are, as commonly done, sorted according to date.

6.2.2 Solving for the coefficients

We now have two ways of calculating all the a_m :

1. Least-squares methods: Following the standard least squares procedure, we differentiate the total error in equation (16) with respect to a_j , a particular coefficient, and set to zero so that error is at a minimum:

$$\frac{\partial e}{\partial a_j} = 0 = 2 \sum_{i=1}^N w_i f_j(h_i) \left(\sum_{m=0}^M a_m f_m(h_i) - F(Q_i) \right), \quad \text{for } j = 0 \dots M. \quad (17)$$

Dividing by 2 and re-arranging the orders of summation:

$$\sum_{m=0}^M a_m \sum_{i=1}^N w_i f_j(h_i) f_m(h_i) = \sum_{i=1}^N w_i F(Q_i) f_j(h_i), \quad \text{for } j = 0 \dots M, \quad (18)$$

thus giving a system of $M + 1$ equations in the $M + 1$ unknowns, the so-called *Normal Equations* in terms of two different types of sum over all the data points. Interpreted in a matrix equation sense, equation (18) is $[A_{jm}] [a_m] = [b_j]$. To evaluate the elements in those matrices, abandoning the usual convention that numbering starts at 1,

For j from 0 to M

For m from 0 to M

$$A_{jm} = \sum_{i=1}^N w_i f_j(h_i) f_m(h_i)$$

$$b_j = \sum_{i=1}^N w_i F(Q_i) f_j(h_i)$$

These equations and the matrix are famously poorly-conditioned unless care is taken to use functions which have some form of orthogonal nature, so that for $j = m$ the sums over the $f_j(h_i) f_m(h_i)$ are rather larger than those for $j \neq m$, and the matrix is diagonally-dominant. If we were to use monomials in stage $f_m = h^m$, as described in books and standards this would certainly not be the case, and this might also explain why global approximation seems not to have been widely adopted.

2. Optimisation methods: These minimise e , usually by black box methods provided by common software such as the Optimisation Solver module in spreadsheets such as Calc in Open Office or Solver in Microsoft Excel, or in a number of other software packages. The methods used might be gradient methods, such as a multi-dimensional Newton's method. It is possible to use the software without a great deal of knowledge of the program details. The methods are powerful, even for nonlinear problems. Our problems here are linear, as the a_m occur linearly in the approximating function (15), and numerical solution does not seem to be difficult.

6.2.3 Choice of basis functions $f_m(h)$

The following approaches were considered, here presented in increasing level of robustness. They have been described in §6.1. Almost all but the first could be considered if optimisation is used, however for least-squares and the normal equations, almost certainly only the last is recommended.

- h^m – the monomials suggested in books and standards. This alternative is potentially fragile, and we will not examine it further.
- $(h - h_{\min})^m$ – usually the datum for h is close enough to h_{\min} so that this is effectively what is used in practice.
- $(\log(h - h_0))^m$ where h_0 is the zero flow point as calculated in §3.2. This will be used only in association with $\log Q$ as the dependent variable.
- y_*^m
- y^m
- Chebyshev polynomials $T_m(y)$

6.2.4 Choice of dependent variable $F(Q)$

We will consider the possibilities

- Q – the actual discharge.
- Q^v – using an arbitrary power of Q that one could choose *a priori*. As justified above, we will adopt a value of $v = 1/2$ and use $Q^{1/2} = \sqrt{Q}$ throughout.
- $\log Q$

6.3 Results

We calculated rating curves for a number of different stations and found that the global approximation method gave good results, in particular when approximating in terms of \sqrt{Q} . Here, we present just some representative results, as they demonstrate problems that can be encountered and overcome, and we make general conclusions, based on all the computations we performed but not reported here. A more comprehensive set of results is given in §9, using formulations and parameters based on our deductions from this section.

Figure 11 shows the results for a rather demanding problem, from figure 2 of Mirza (2003), for the River Ganges in the year 1992. We scanned the data from a small figure, so our accuracy of reproduction of the data may be wanting. It does, however, provide a very interesting test for the theory, as it contains some detailed structure. It is clear that over most of the range of stages, the global polynomial approximation has worked very well, describing apparent small-scale fluctuations in the data, although it was necessary to go to $M = 10$ to do this satisfactorily. However, as we have often found to be the case using the actual discharges Q_i , at the bottom end there are significant oscillations in the approximation. Throughout all our results it should be remembered that we are obtaining a polynomial for Q or \sqrt{Q} , plotted horizontally so that deviations and considerations of quality of fit are actually horizontal ones. Using $\sqrt{Q_i}$ instead of Q_i , the method worked very well, as shown. It did so throughout all of our examples.

Each curve in the figure shows the identical results of four approximations, first using as basis functions the monomials $f_m(h_i) = (h_i - h_{\min})^m$ in terms of the physical stage relative to the minimum stage value. In this case, with $h_{\min} \approx 6$ m the subtraction is highly necessary. The next approximation was using the vertical variable y_* scaled to $[0, +1]$, the next was the vertical variable y scaled to $[-1, +1]$ and the fourth using Chebyshev polynomials, $f_m(h_i) = T_m(y_i)$, a process that should have been much more numerically robust. To our surprise we found that the plotted results even from the questionable basis functions coincided to more than 5 significant figures. However, they hide something that has important implications for the application of global approximation, and which we have already foreseen from the appearance of the different functions. Figure 12 shows a particular result – showing how the

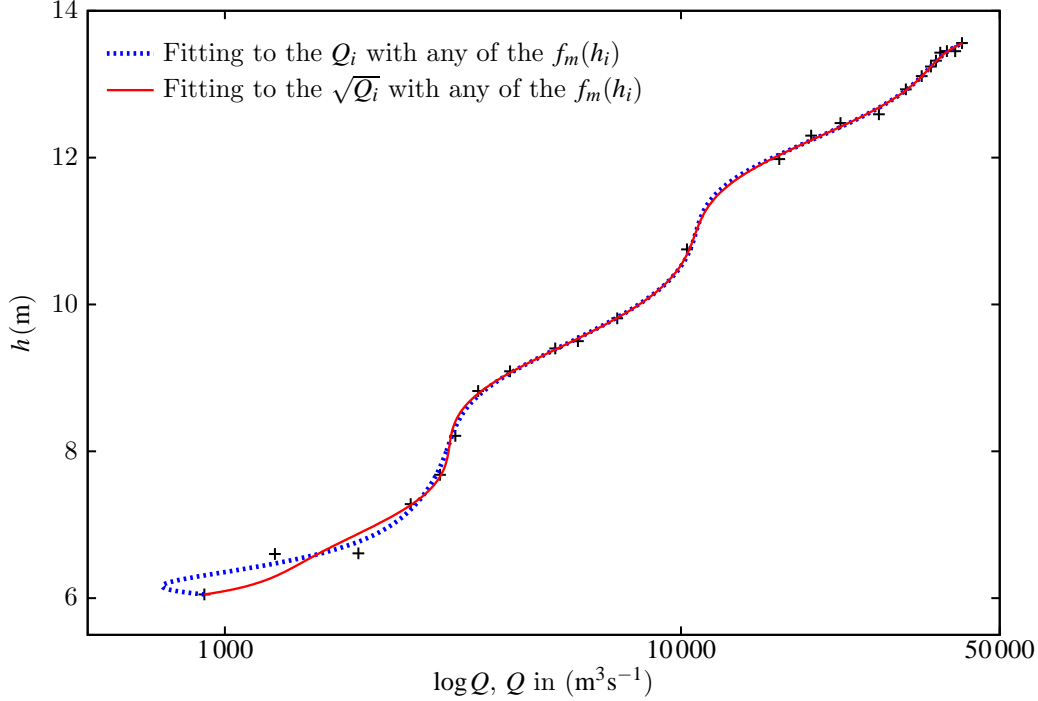


Figure 11: River Ganges in 1992, from Mirza (2003) – semi-log plot

terms in different approximating polynomials contributed to the result for the discharge at the maximum stage $h_{\max} = 13.56\text{m}$ shown in Figure 11. We wanted to examine how the sum of the contributions $\sum_{m=0}^M a_m f_m(h_{\max})$ converged in m . As some contributions are negative we have actually plotted the absolute values $|a_m f_m(h_{\max})|/Q_{\max}$, dividing by the maximum discharge in the data set $Q_{\max} \approx 41300\text{m}^3\text{s}^{-1}$ to scale the results (we considered the case for discharge Q without taking the square root). We expect the sum to be close to 1 as the approximations pass close to such an isolated point. As individual terms were surprisingly large, we have used a logarithmic scale.

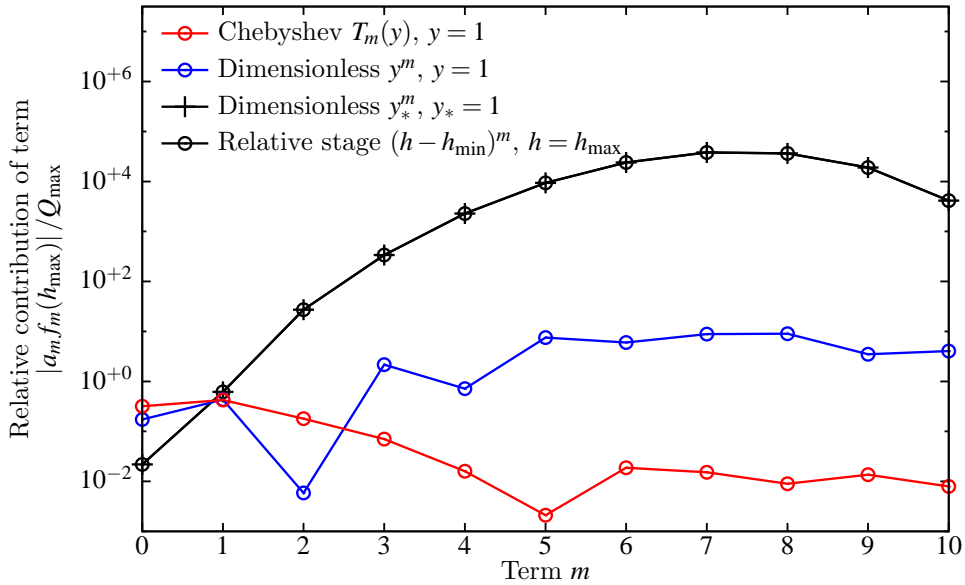


Figure 12: The relative contributions of terms in polynomials with different basis functions, for the single case of calculating an approximation to Q using the maximum stage at the right of Figure 11.

The first results we examine are those from the use of Chebyshev polynomials $T_m(y_i)$ on the interval

$[-1, +1]$, obtained by both numerical solution of the normal equations and optimisation. It was encouraging (and to be frank, surprising, as the solution process in the two cases is completely different) that the results from the two methods agreed closely – to at least five significant figures. This gives us confidence in results from both. What is equally satisfying but more important is that the contributions, shown by the bottom red line and circles, are all small, as one would want, with typical contributions $\approx 10^{-2}$, rather less than approximately 1 to which they should sum, and, they become smaller as m increases, showing that the process is converging.

Now, however, if we relax our standards, and instead of $T_m(y)$ we simply use the monomials y^m with y defined on $[-1, +1]$ we find that the individual contributions are now not approximately 10^{-2} , but approximately $10^0 = 1$, so that we have 10 contributions, each of approximate magnitude ± 1 summing to about 1, which is not so desirable from a computational point of view. The sum, however, the calculated value of discharge, was the same as that from using Chebyshev polynomials, to 10 significant figures.

The non-converging magnitude of those contributions pales into insignificance, however, when we consider the results from the top curve, from polynomials in $(h - h_{\min})$ and in the scaled variable y_* in $[0, 1]$. To the author's surprise, the scaling has had no effect, but to the author's much greater surprise, the individual contributions are actually huge, of order 10^4 (oscillating in sign of course). In Figure 11 we were not to know that we were looking at the results of a sum, each term with a magnitude of plus or minus 10000 times that of the final sum!

This would be a problem for widespread application. Within a particular computer program, as we found, the results were good, but if one were reporting results or transferring between programs, unless using Chebyshev polynomials, the greatest care would have to be taken to specify coefficients to sufficient accuracy. In any case, it is ugly to have such sums with hugely oscillating contributions.

Conclusion 1 concerning global polynomial approximation: *in spite of the warnings above, it is possible to use polynomials made up of simple monomials in $h - h_{\min}$, although it is better to scale using y_* in $[0, 1]$ or y in $[-1, 1]$ so that the coefficients do not become very small for higher degree terms, in case one has to pass them between persons or programs. However, it is highly desirable, if not essential, to use Chebyshev polynomials $T_m(y)$ as basis functions, so that individual contributions in the polynomials do not oscillate unreasonably, in case low-accuracy arithmetic is being used or results passed between persons or programs.*

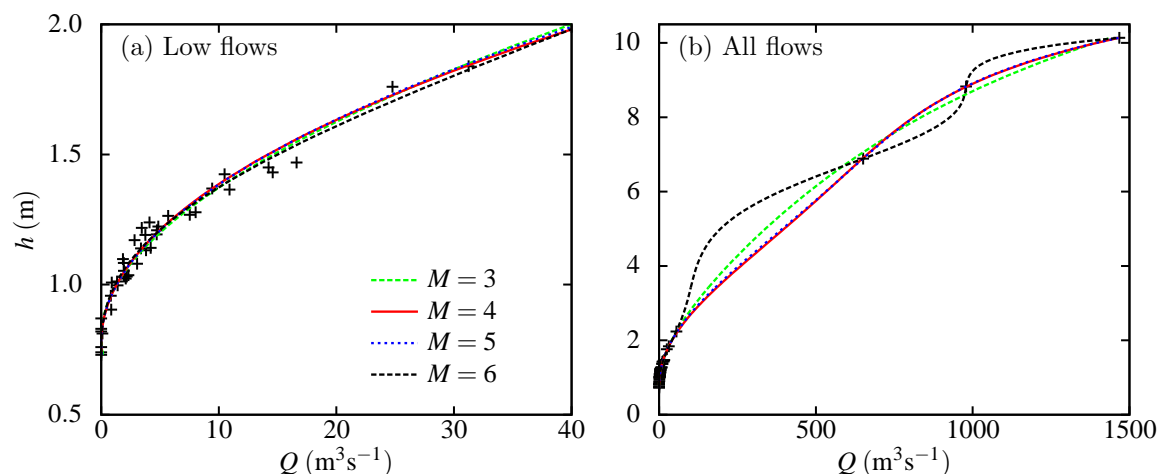


Figure 13: Typical results for variation with degree M using global approximation, here for data with gaps from the Gwydir River at Pallamallawa (see p.38).

Figure 13 is presented to show typical behaviour of global polynomial approximation with respect to the degree M of the polynomial. The problem is actually a quite difficult one, with some large gaps

in the data. An unusual aspect is that amongst the data points are some of zero flow. This means we could not use a logarithmic scale to plot the discharge. Where there is a high density of data points, at the low end, shown in part (a), and variation is smooth and regular, the approximation works well, even for low-degree polynomials. Part (b) of the figure shows all flows, including data with large gaps between points. It can be seen that even for as low a polynomial degree as a cubic, $M = 3$, an adequate smoothing approximation is obtained, which might well be accurate enough (the author knows that flow measurements for some of these highest flow points were obtained by boat, floating over the surrounding flooded agricultural fields!). For $M = 4$ and 5 the results are almost coincident, and reveal a desirable aspect of the approximation, that the approximation passes close to isolated points. However, if the degree of approximation is too great for the spacing of data points, such as for $M = 6$ shown, the polynomial can fluctuate wildly. This happened for all global approximations to rating problems we considered – sooner or later a level of approximation is reached where large fluctuations occur. Unfortunately one can never predict when this will happen. We have already seen in Figure 11 that, provided enough points are uniformly distributed, high degrees of approximation are possible.

Conclusion 2 concerning global polynomial approximation: *beyond a certain degree of approximation, large fluctuations occur, becoming worse for higher degrees. This is especially so if there are large gaps between data points.*

We have considered trying to quantify this so that one could determine an appropriate degree of approximation without having to examine results of several levels visually. We considered two numerical quantities that could be calculated and monitored:

1. The total error e as defined in equation (16) and calculated in the course of optimisation, the sum of the squares of the errors over all the data points. It would tell us how good the approximation is over all the data points. It would decrease with increasing M as the approximation gets better, but it would not tell us anything about fluctuations between points.
2. The "roughness" or total mean square curvature (or just second derivative) of the polynomial would measure the amount of fluctuations. If the polynomial is written $P(h)$ ($= \sum_{m=0}^M a_m f_m(h)$), this quantity would be

$$\int_{h_{\min}}^{h_{\max}} \left(\frac{d^2 P}{dh^2} \right)^2 dh.$$

For a smooth approximation this would be a certain value commensurate with the data, while with fluctuations, it would be larger. Beyond a certain degree, as approximation passed beyond the limiting smooth value, it would suddenly become larger, which would be an indication of the fluctuations.

These two quantities are different, one measuring goodness of fit, decreasing with M , the other measuring level of fluctuations, possibly remaining roughly constant, until a sudden increase at a particular value of M . They could be combined, but it is not clear how, and at this stage we regrettably conclude that we do not know how to identify the limiting degree beyond which the polynomial will show unacceptable fluctuations, other than by trial and error visually.

We now consider the results in Figure 14, selected from all the computational examples we considered to show more typical aspects – and problems – obtained using global polynomial approximation. Part (a) of the figure is for low flows, part (b) for all the flows, both using natural scales. The blue dotted line is for all the polynomial fits to the actual discharge Q_i , whether using the natural relative stage $(h - h_{\min})^m$, the scaled monomial y_*^m and y^m , or the Chebyshev polynomials $T_m(y)$. Again we find that our warnings about being careful to scale and even better to use the quasi-orthogonal polynomials have apparently not really been justified (however the details of convergence of the polynomials still showed Chebyshev to be best). Part (a) shows a characteristic feature of all our results that approximating the actual discharges Q_i is likely to give oscillations at the ends of the range. Results for $M = 7$ were even wilder. Even for this relatively smooth data set, the method is at the limit of its applicability.

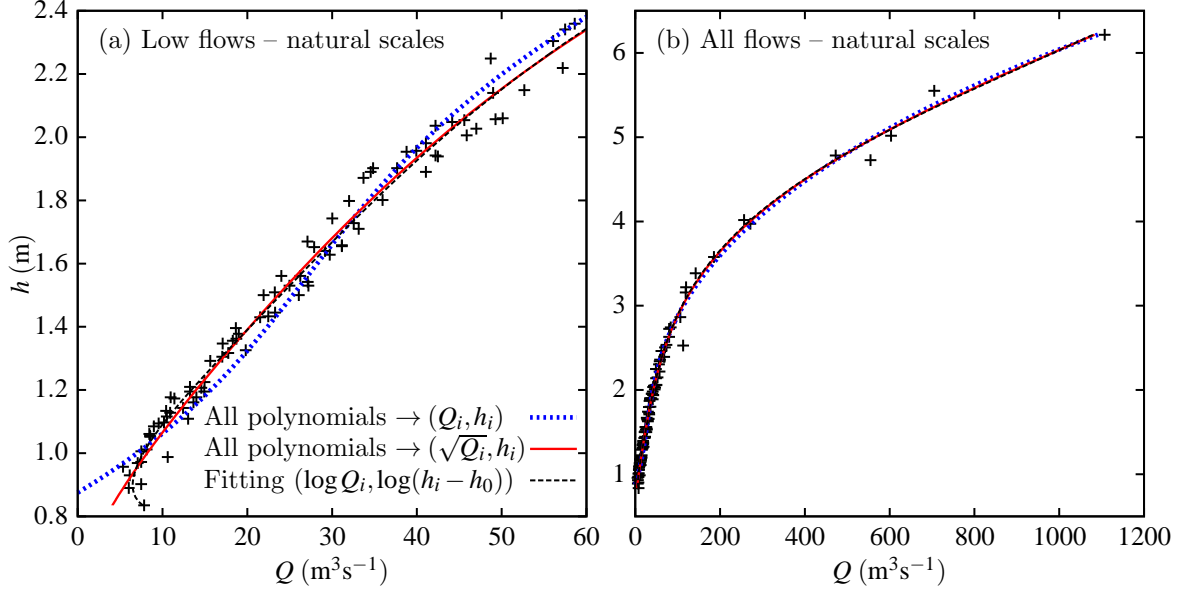


Figure 14: Comparison of global polynomial results for the approximation of different functions of Q , all for degree $M = 6$. Data: USGS Station 02361500 on the Choctawhatchee River, AL, USA, 2000-12-07 to 2015-05-22.

The results for all polynomial fits to the $\sqrt{Q_i}$ shown by the solid red line, however, describe it well.

The values obtained for the exponent for the two power function fits for this station to points $Q_i < 15$ and $Q_i < 30$ shown in Figure 1 were $\nu = 0.68$ and $\nu = 0.40$, respectively, with visually almost the same fit to the points. It can be seen that the value $\nu = 0.5$ that we chose as a constant representative value, leading to fits using $Q_i^{1/2} = \sqrt{Q_i}$, was appropriate in this case at least. In fact, it always seemed a robust approximation.

The third approximation shown on the figure, shown by a black dashed line is one with origins in the power function of §3. Here a polynomial fit to the data points $(\log Q_i, \log(h_i - h_0))$ was made using basis functions $(\log(h - h_0))^m$. It was necessary to calculate h_0 initially using the method described in §3.2 for points with $Q < 30$. The value of $h_0 = 0.547\text{m}$ used was the higher of the two values shown in Figure 1. The global approximation here approximates the data points very well for these low flows. In fact, possibly too well, turning around as the more scattered lowest points are encountered, faithfully but probably not reasonably, following them. In the approximating space with logarithmic scales, the low flow end is expanded, and the approximation tries hard – and successfully – to approximate in that region.

Figure 14(b) shows the behaviour of the approximations for larger flows, where results for Q_i and $\sqrt{Q_i}$ are almost coincident. The curve from the $(\log Q_i, \log(h_i - h_0))$ scheme gave slightly different results, although these are barely visible here. The natural and square root schemes usually worked quite well, while the double logarithmic scheme often gave wild results for high flows, where the region is contracted, and variation is apparently more rapid, making approximation more difficult.

Conclusion 3 concerning global polynomial approximation: *approximating the square root of the discharges $\sqrt{Q_i}$ has been always found here to be more accurate and robust than using the Q_i . The reason is probably a combination of two factors. One is that for small discharge the data points often vary roughly like Q^ν , where $\nu \approx 0.5$, so that approximating the Q^ν themselves is computationally simpler. The second is that the magnitude of variation between the minimum and maximum discharges is much less (instead of a factor of 10^4 it would be 10^2), so that the global approximation does not have to work so hard. The use of logarithms for both discharge and stage also accomplishes this, but gives too much attention to the low flow end and too little to the high flow end, so that fluctuations there are more likely.*

6.4 Conclusions concerning global polynomial approximation

The method works well, especially if the square roots of the discharges $\sqrt{Q_i}$ are approximated. Good results seem to be able to be plotted even just using monomials in the relative stage $h - h_{\min}$, despite the warnings above. However, to maintain and use the results in working software it would be better to use Chebyshev polynomials, when the coefficients in the polynomials are very much smaller, making their use less sensitive to errors in numerical processing. In §9 we will see that for seven stations we show good results, using the degree of the polynomial that we found best in each case. That degree is never clear *a priori*, and it seems that trial and error and visual evaluation is necessary to establish it. The method is fragile, however, for usually taking a polynomial one degree higher gave wild fluctuating results.

To obtain a method which is less-sensitive to that computational parameter we will describe the development of another method for approximating rating data which is indeed more flexible and robust, but first we mention another known method.

7 Local approximation

Summary: The "LOWESS Smoothing Method" is a local approximation method. It is briefly described. A better method will be developed in the next section.

The acronym LOWESS or LOESS is awkward, it stands for "LOcally Weighted Scatterplot Smoothing". We call it here local approximation: it fits simple local approximations to data using least squares, giving more weight to points near the point whose response is being estimated and less weight to points further away. The procedure seems to be arbitrary and complicated. User-specified input to the procedure determines how much of the data is used to fit each local polynomial. The smoothing parameter is used to determine the proportion of data used in each fit and controls the flexibility of the approximation function. Large values produce the smoothest functions that wiggle the least in response to fluctuations in the data. The smaller the parameter is, the closer the regression function will conform to the data. The local polynomials fit to each subset of the data are almost always of first or second degree; that is, either locally linear (in the straight line sense) or locally quadratic. Higher-degree polynomials would work in theory, but yield models that are not really in the spirit of the method. It is based on the ideas that any function can be well approximated in a small neighbourhood by a low-order polynomial and that simple models can be easily fitted to data. A smoothing parameter value has to be provided as well, of course, as the degree of the polynomials.

That is all we will do here to consider the method. Before the present author discovered its existence he had devised a method, later found to be well-known, which shares many of the characteristics of the Local Approximation method we have just described. However it seems to be simpler and to have a readily-understandable physical significance. There are no arbitrary smoothing parameters. We will now describe it and apply it to the approximation of rating curve data, where it will be seen to be quite powerful.

8 Approximating splines

Summary: Even though the global approximation method gave good results for all problems we considered, in the background are the numerical problems we described, including poor convergence of series and unacceptable fluctuations after a certain degree. Such problems can be overcome by using splines. Here we develop a theory and method for using low-degree spline functions to approximate rating data in a least-squares sense. For many problems very little operator involvement is necessary, possibly just specifying the number

of spline intervals to be used. For others, some human interaction might be needed for fine tuning, by specifying the stages of the knots defining the splines; the discharges at those points are calculated automatically.

8.1 Introduction

We have seen how global interpolation and approximation have some problems. For interpolation it is well-known that, on the other hand, piecewise-continuous polynomial splines have advantages, especially where there are regions of local rapid variation. We will apply them to problems of approximation. For such problems there are computer programs for *Smoothing Splines*, which approximate data in a piecewise continuous sense, for example de Boor (2001). A problem, however, is that no two data points can have the same value of the independent variable (h in our case). In rating data, it is quite possible that there will be at least two points with the same measured value of stage, thus ruling out this approach.

An obvious alternative method is to use splines with a finite specifiable number of knot points independent of the data points, across which the function and some of its derivatives will be continuous, but the actual values of the functions at the knot points will be determined by least-squares approximation. The knot points will, in general, be spaced more closely where the data varies more rapidly. It is to be expected that with the ability of the chosen quadratic or cubic polynomials to describe variation across an interval, that not many such intervals will be necessary. We will call this method that of "approximating splines", emphasising what it does and what method it uses.

After we developed this method, we found that it was of wide general applicability to any scattered data. We have made a paper (Fenton 2015a) and computer program available at <http://johndfenton.com/Approximating-splines/>. Here we repeat some of that theory, sometimes word-for-word, with special reference to rating curves.

Prosdocimi (2015, Personal Communication) informed us that such methods are well known in the statistics literature, and alerted us to several sources, including the book by Wood (2006). She wrote "In all the statistical literature though, splines are used with the aim of modelling the central part of the distribution data, so there will be possibly less attention than you would like to fitting data well in the tails". She went on to cite the paper by Eilers & Marx (1996) which she described as the standard reference paper to penalised splines "which are largely used nowadays when doing nonparametric regression. The idea of penalised splines is that rather than choosing carefully how many knots to have and where to place them, one takes a large basis and then penalise the likelihood to make sure to avoid over-fitting". We have noted these remarks, but will continue to present the relatively simple and maybe naive theory we had developed, which works well even for the ends of the data. We will leave the second remark, concerning knot placement, for attention at a later date.

8.2 Theory and computational method

Consider the scattered data such as shown in Figure 15 comprising a number of data pairs (h_i, Q_i^v) with $i = 1, 2, \dots$. We generalise our use of Q and \sqrt{Q} by just writing Q^v generally, where $v = 1$ or $1/2$. In practice one might experiment with other values of v for each station, but we have always found $v = 1/2$ to be most satisfactory. Let the data interval $[h_{\min}, h_{\max}]$ be subdivided into J intervals by $J + 1$ points, where $h = h_j$ for $j = 1 \dots J + 1$. As in our description of interpolation, our notation is slightly ambiguous, but clearer than alternatives: we use h_i for data points, h_j (and h_{j+1} etc.) for interpolation points and here, knot points). We expect that the h_j will be placed roughly in accordance with the rapidity of variation of the data. They must be separate and ordered, such that $h_{j+1} > h_j$ for all the j . The points are to be approximated by a number of low-degree polynomials such as quadratics or cubics, $P_1(h - h_1)$, $P_2(h - h_2)$, \dots , $P_J(h - h_J)$, where the number indicates the interval over which the polynomial $P_j(h - h_j)$

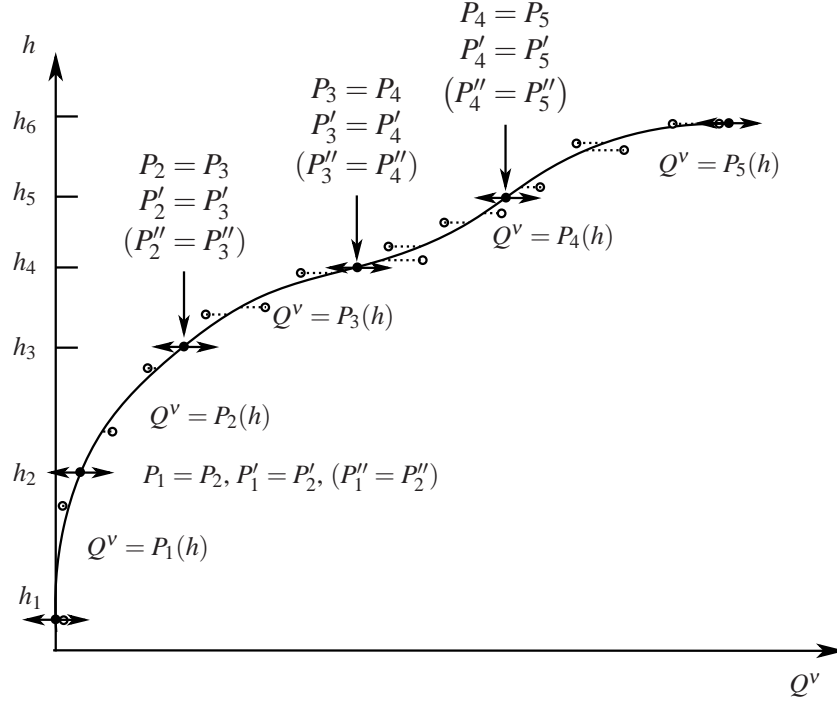


Figure 15: Spline approximation of scattered rating points

is valid, between knot points at $h = h_j$ and $h = h_{j+1}$. Over each interval we have a polynomial of degree M , such that at over the interval j , $h_j \leq h \leq h_{j+1}$:

$$P_j(h - h_j) = c_{j,0} + c_{j,1}(h - h_j) + c_{j,2}(h - h_j)^2 + \dots = \sum_{m=0}^M c_{j,m}(h - h_j)^m, \quad (19)$$

where in this case, the degree M is expected to be only 2 or 3, giving quadratic or cubic functions, much the same as in polynomial spline interpolation. It is not necessary to scale the h as we recommended for global polynomial approximation in §6, as it only appears as the local shifted value $h - h_j$ and the degree of the polynomial is low.

Figure 15 shows the essential nature of the approximation, showing separate polynomials in between each pair of knot points, satisfying two or three continuity conditions at each knot point, whose h_j may be set as data or determined automatically. The knot points are shown with double arrows suggesting how the program varies the value of their Q_j^v so that the sum of the squares of the differences between data points and functions, shown by dotted lines, are minimised. In the approach here, only those data points falling within a certain interval contribute to the total sum using the function for that interval. In this sense the method is local; where it becomes partly global is in the continuity conditions at the knot points. Now we specify mathematically the conditions that we have tried to show graphically.

The spline nature of our approximation now requires us to satisfy across each interior knot the continuity of function value plus all derivatives up to $M - 1$. Hence for all $j = 2 \dots J$ we have

$$P_{j-1}(h_j) = P_j(h_j), \quad (20a)$$

$$P'_{j-1}(h_j) = P'_j(h_j), \quad \text{and if } M = 3, \quad (20b)$$

$$P''_{j-1}(h_j) = P''_j(h_j). \quad (20c)$$

where the primes indicate differentiation with respect to h . From equation (19) at left and right of each

interior knot point, and using $\delta_j = h_{j+1} - h_j$ for the interval length, the continuity conditions become

$$c_{j+1,0} = \sum_{m=0}^M c_{j,m} \delta_j^m = c_{j,0} + c_{j,1} \delta_j + \dots + c_{j,M} \delta_j^M, \quad (21a)$$

$$c_{j+1,1} = \sum_{m=1}^M m c_{j,m} \delta_j^{m-1} = c_{j,1} + \dots + M c_{j,M} \delta_j^{M-1}, \quad (21b)$$

$$\text{and if } M = 3, \quad 2c_{j+1,2} = \sum_{m=2}^3 m(m-1) c_{j,m} \delta_j^{m-2} = 2c_{j,2} + 6c_{j,3} \delta_j. \quad (21c)$$

Now, unlike Interpolating or Smoothing Splines, we do not use the conditions that the knot points are data points. Neither do we require, like Interpolating Splines, that each spline passes through the corresponding data point $P_j(h_j) = Q_j^v$. Instead, we seek to approximate the data points such that the sum of the squares of the errors over all the points is minimised. We write this as a sum over all the intervals of all the contributory points in each interval, and where each data point is assigned a weight of w_i :

$$e = \sum_{j=1}^J \sum_{i \in I_j} w_i (P_j(h_i - h_j) - Q_i^v)^2, \quad (22)$$

where we have used the mathematical notation $i \in I_j$ where I_j is the set of points which are in interval j , $I_j = \{i : h_j \leq h_i < h_{j+1}\}$ which simply means taking all the points i which are in interval j . This means that the contribution to the total error e of a data point is only given by the spline function on the interval in which it falls. It will affect the overall result by the continuity conditions at the ends of that interval.

We write e using the polynomial as given in equation (19) as

$$e = \sum_{j=1}^J \sum_{i \in I_j} w_i \left(\sum_{m=0}^M c_{j,m} (h_i - h_j)^m - Q_i^v \right)^2. \quad (23)$$

Convenient aspects of this formulation, as for global polynomial approximation, are:

- The data points h_i can be in any order
- There can be multiple data points with the same h_i , and
- The weights w_i can be assigned arbitrarily so as to attach less or more or even zero importance to a point, for example to filter according to time period, or, with a large weight, to force the spline to go through or near a point.

We calculate how many unknowns we have. For each interval $j = 1 \dots J$ we have $M + 1$ values of the $c_{j,m}$ giving $J(M + 1)$ unknowns or degrees of freedom. However, equations (21) enable us to eliminate M unknowns at each interior knot point $j = 1 \dots J - 1$ giving a total of $M(J - 1)$ such unknowns, so that the net number of unknowns is $J + M$, which is not large. The unknowns are all the $M + 1$ coefficients in the first interval: $c_{1,0}, \dots, c_{1,M}$ plus the M th degree coefficients at each of the $J - 1$ internal points $c_{2,M}, \dots, c_{J,M}$.

The problem is now to find all the $c_{j,m}$ such that e is minimised. As in §6.2.2 for global polynomial approximation, we have two ways of calculating them. The second method is rather simpler and easily implemented.

1. Least-squares methods: For global approximation such as a polynomial, the equations are famously poorly conditioned, and numerical solution can be quite difficult and not so accurate, but which we seemed to obviate using Chebyshev polynomials. The spline formulation would lead to a diagonally-dominant matrix with smaller blocks of elements centred on the diagonal. This would be more robust. Differentiating equation (23) with respect to a coefficient $c_{k,l}$, and setting to zero

for an extremum, only a single value of $j = k$ in the outside summation contributes, and so we have

$$\frac{\partial e}{\partial c_{k,l}} = 2 \sum_{i \in I_k} w_i (h_i - h_k)^l \left(\sum_{m=0}^M c_{k,m} (h_i - h_k)^m - Q_i^y \right) = 0,$$

for $k = 1 \dots J$ and $l = 0 \dots M$. Dividing by 2, re-writing and reverting to using j instead of k :

$$\sum_{m=0}^M c_{j,m} \sum_{i \in I_j} w_i (h_i - h_j)^{l+m} = \sum_{i \in I_j} w_i Q_i^y (h_i - h_j)^l,$$

for each of the j and l , giving a system of $J(M+1)$ linear equations, with a similar nature to the normal equations (18), to which must be added the compatibility conditions (21). If we were to interpret this in a matrix sense, for each interval j there are $M+1$ rows from the error minimisation, each involving elements pertaining just to the four unknowns for that interval, plus $M-1$ rows for the compatibility conditions, also involving unknowns from the next interval.

2. Optimisation methods: again we minimise e in equation (23) using common software. There is a choice, whether to actually eliminate variables using equations (21) thereby guaranteeing continuity of all derivatives up to the $M-1$ at each interior point, or, whether to add the square of the equations to e in a weighted sense, and solving the whole numerically. Thus, for example, adding terms to e in equation (23) such as

$$e = \dots + \lambda_0 (c_{j,0} + c_{j,1} \delta_j + \dots - c_{j+1,0})^2 + \lambda_1 (c_{j,1} + 2c_{j,1} \delta_j + \dots - c_{j+1,1})^2 + \dots,$$

in which the $\lambda_0, \lambda_1, \dots$ are weights for these compatibility conditions. This introduces a certain arbitrariness, and we found that it did not work so well. In all the results presented in this report, the author chose the former approach, to eliminate the variables and have exact continuity up to the $M-1$ derivative across each knot.

Here we adopt the simpler and more modern method, where we just use optimisation software to minimise e and determine the values of the coefficients. Such software is widely available, including the SOLVER module in EXCEL spreadsheets, however in this case, the programming is rather more complicated than for global approximation because of the decision-making as to into which interval a point falls.

8.3 Number of intervals and placement of knots

One has to have enough intervals so as to be able to describe the local variation adequately, but with enough data points in each interval to define the local quadratic or cubic. The method does a very good job of approximating whatever it is given, but sometimes this is too good, for example when in one interval there might be as few as 2 or 3 data points, the local quadratic/cubic obligingly tries to interpolate them, passing through or close to each point. This is not what one wants with scattered data, so that in the lower and middle parts of the data it is desirable to have a minimum number of points in each interval. Near the high-flow ends of the data sets, where there might be isolated rating points, it is not such a problem: on the contrary it is a benefit, as the results tend to agree closely with the data, which in this case is a pleasant property. It is possible to have an interval without any data points, sometimes convenient for high flow regions.

It is recommended for the first application of approximating splines to a problem that the default option be chosen, where the program places the knots automatically such that there are equal numbers of data points in each interval. Then, visual examination of the results might suggest the clustering of knots in regions of rapid variation, or the expansion where there are few data points. A knots data file could be prepared with values of the h_j modified by hand. In typical applications described below, values of $J = 3 - 10$ intervals have been found to be satisfactory for all problems considered.

8.4 Solving for coefficients

The author has implemented the method in three different software environments: using an inbuilt function in the MAPLE language and environment; in the C language using Powell's method for optimisation of a function (§10.5, Press, Teukolsky, Vetterling & Flannery 1992); and in the SOLVER module in the spreadsheet EXCEL. In all cases the method and solutions obtained seemed to be quite robust.

8.5 Performance

The program has worked surprisingly well in all the examples the author has considered. It has never failed badly. However, adjustment of the knot positions may be necessary to describe regions of high curvature or to avoid having too few points in an interval. Generally the results are better if a value of $\nu = 1/2$ is used, that is, approximating in (h, \sqrt{Q}) space, as was found for global polynomial approximation.

8.6 Degree of splines: quadratic or cubic

It does not seem to matter very much whether quadratic or cubic splines are used. The method is quite robust, with a minimal tendency for oscillations to develop. One may as well use quadratic splines, as there are fewer unknowns and computation time is shorter, even if that was less than 1 s in the examples we considered.

9 Results

***Summary:** Results of the global approximation method and of the approximating spline method are presented for seven gauging stations. They are highly satisfactory. If anything, the splines rather than global polynomials are to be preferred. Using both would be a sensible practice.*

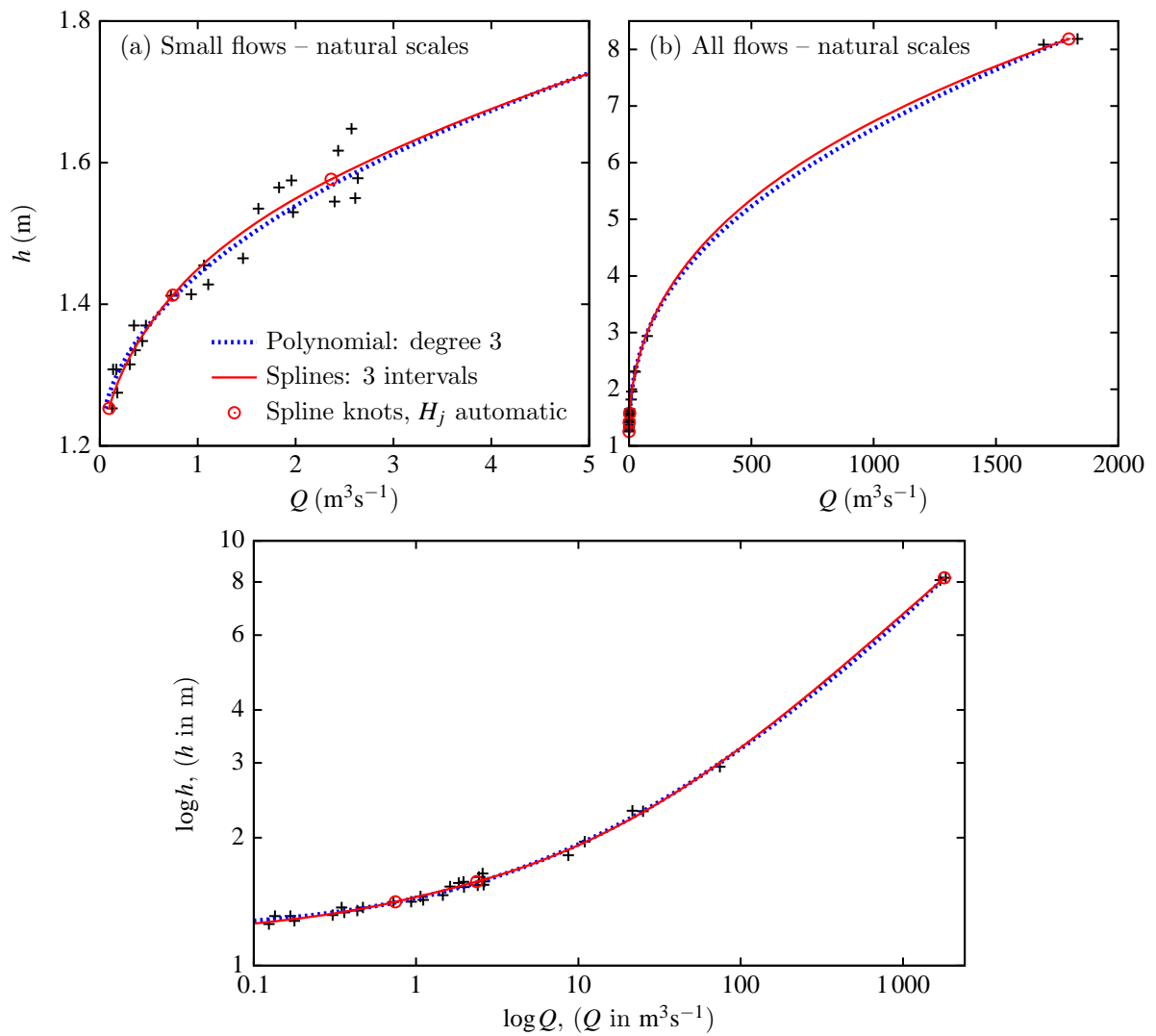
We now examine the performance of the method for a number of different gauging stations in Australia, the Brahmaputra River, Ganges River, and in the USA. Wherever necessary we have converted to *SI* units.

In each case we present two approximations, one global, the other piecewise-local splines. All have been obtained using $\nu = 1/2$, namely, approximating $Q^{1/2} = \sqrt{Q_i}$. In all cases the best possible global approximation has been obtained by varying M , the value of which is shown on each figure. We tried to use the smallest value compatible with accuracy. Often increasing M by a single degree beyond that led to unacceptable fluctuations.

For the approximating splines, everywhere quadratic splines $M = 2$ were used. In some cases, automatic allocation of knot points was used; in others, especially for large flows where there were few data points, we allocated the knot points by hand.

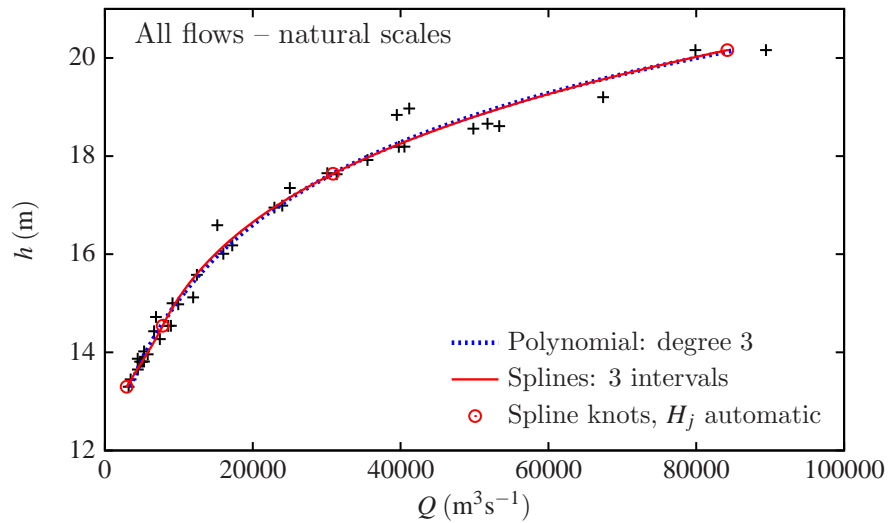
Most figures have three parts, (a) the low flow end, (b) all the flows, both using natural scales, and (c) using log-log scales. We slightly disapprove of the latter, but they do show more clearly the behaviour at the low-flow end.

9.1 Avon River at Stratford, Vic. Australia, site number 225201A, 2012-06-05
 – 2015-04-24



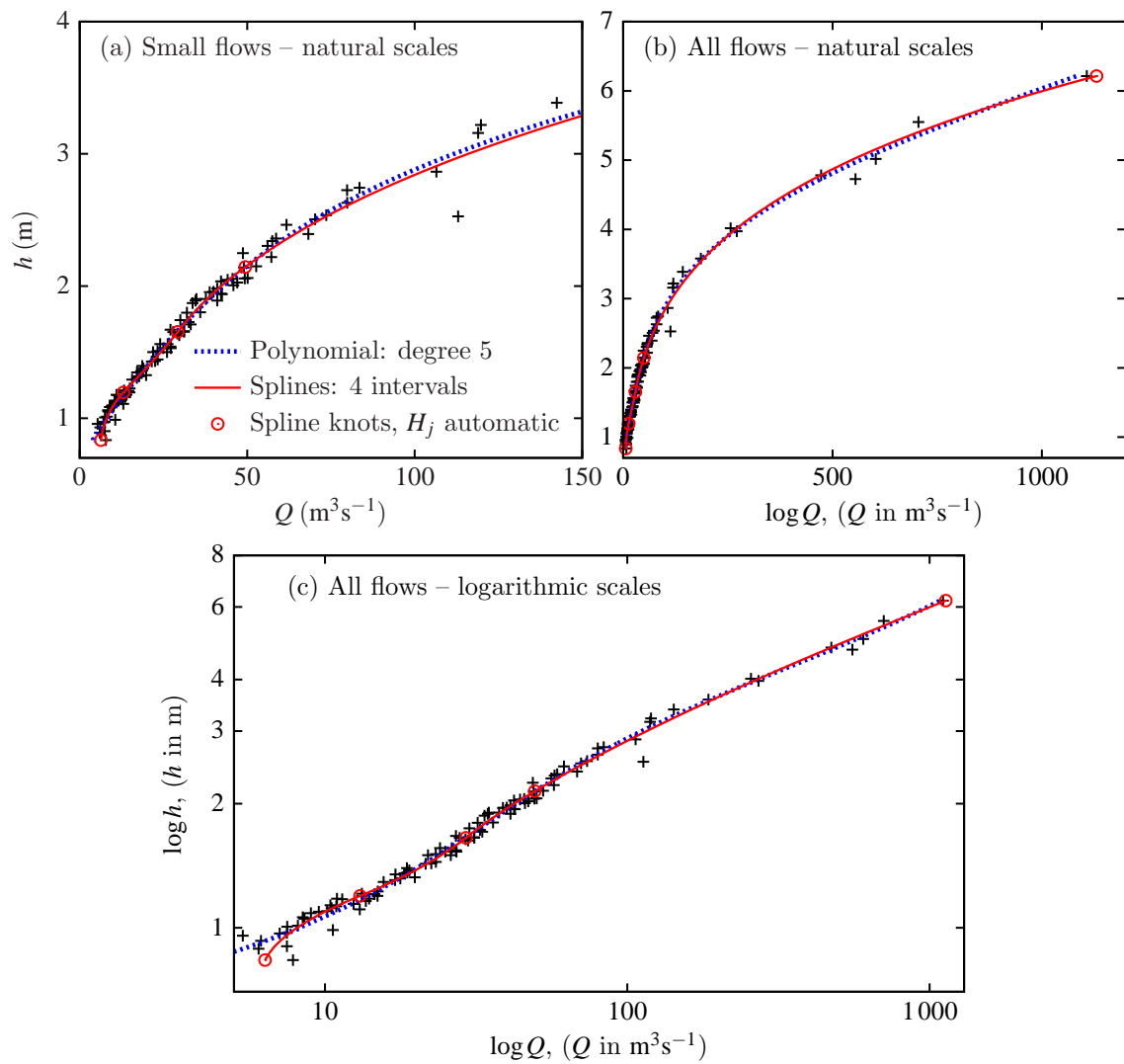
This is an unusual data set with a large gap in it. The spline intervals were allocated automatically to give the same numbers of data points in each. The global polynomial of low degree, $M = 3$, and the quadratic splines, have worked well, giving a plausible bridging of the gap. For higher degree polynomials, both the isolated data points for the highest flows were almost interpolated, so that the polynomial results deviated from those of the splines. We cannot say which was right.

9.2 Brahmaputra River 1992



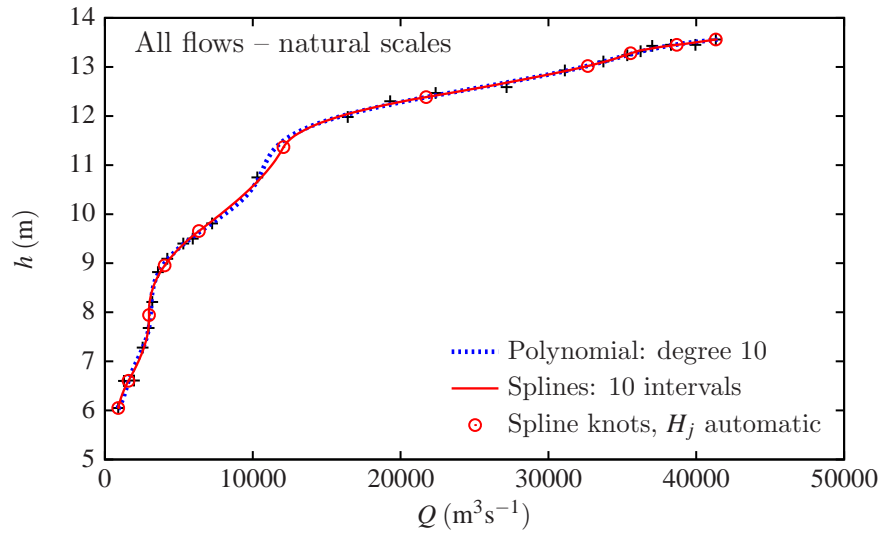
The data was taken from Mirza (2003) by digitising a fairly small figure, so our accuracy might be questionable. Both methods handle this fairly simple problem quite well, again with low levels of approximation, although the polynomial shows a characteristic fluctuation at the low flow end, not visible on this figure. It might be thought that neither method has treated the two points at $h \approx 19\text{m}$, $Q \approx 40000\text{m}^3\text{s}^{-1}$ well, but it should be remembered that our approximation is for Q , plotted horizontally, and so it actually performs quite well in passing between those two points and the three or four points on the other side of the curves with roughly the same stage.

9.3 Choctawhatchee River near Bellwood AL, USA, USGS Station 02361500, 2000-12-07 to 2015-05-22



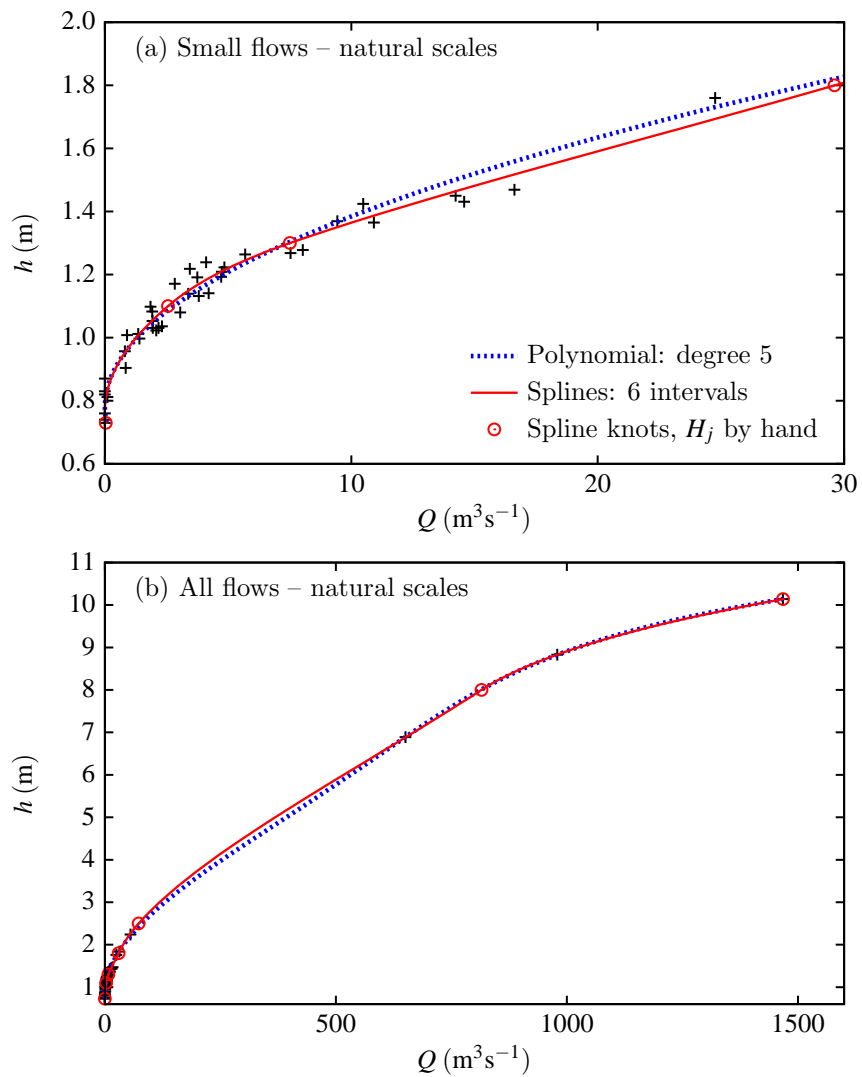
The only real problem for approximation seems to be for the lowest flows, where there is a quite characteristic scatter of points, made more obvious in part (c), the log-log plot. The splines seem to have picked this up. It is surprising and heartening that such a non-trivial structure can be described with just quadratic equations over few intervals.

9.4 Ganges River 1992



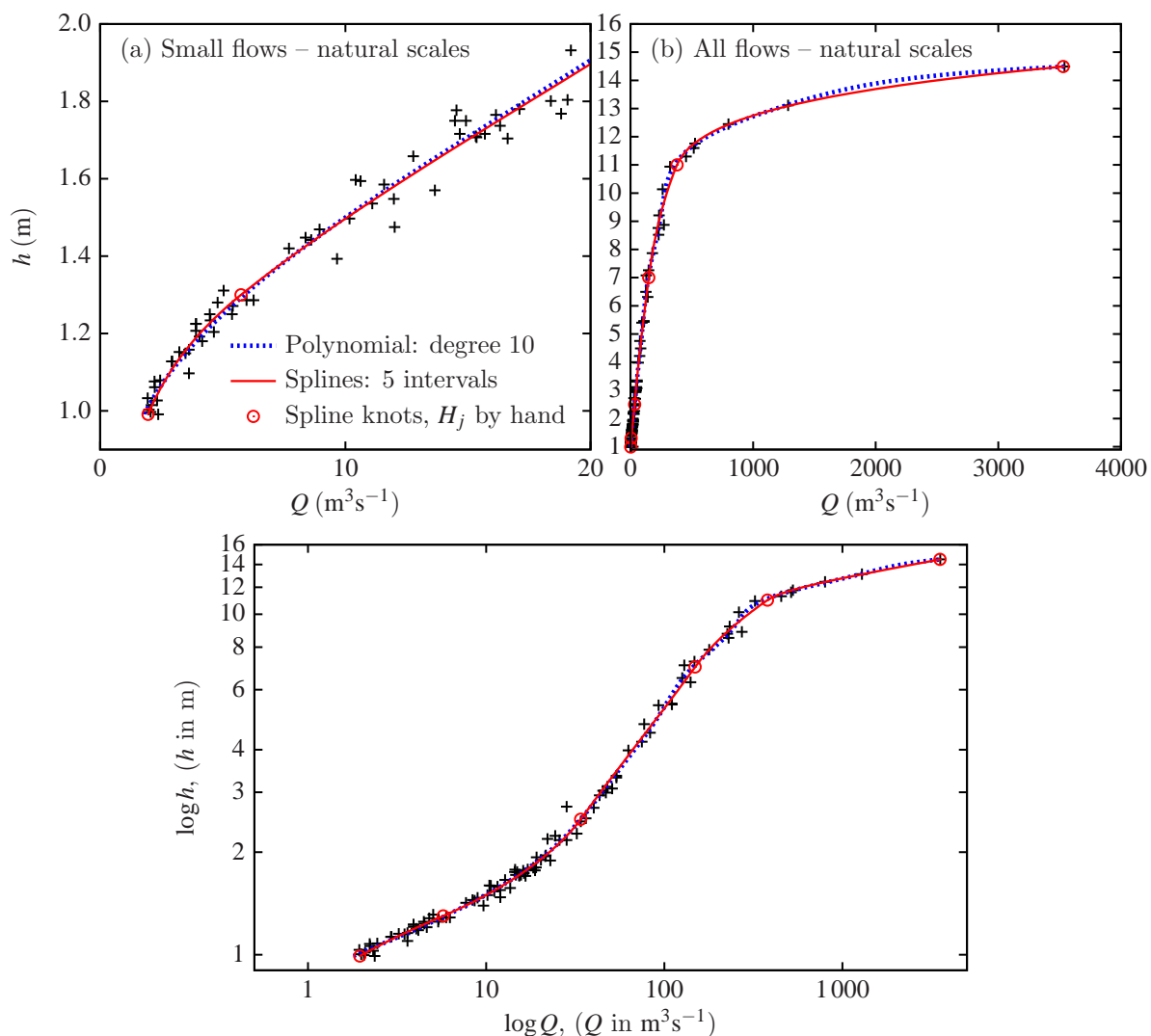
The data was also digitised from a small figure in Mirza (2003). This is a demanding problem, if one accepts that the fine structure ("lumps") in the data are real. We have chosen to do that as a test of the model. Both polynomial and splines, with high levels of approximation, perform well in describing the complicated variation. The polynomial can do this because the data points are well-distributed. The splines have performed well, even if the placement of the knots has yet again been performed automatically, and there are actually few data points in each interval. If one knew the reliability of the data, it might be possible to relax and use a smoother approximation, with a polynomial of lesser degree and/or splines with fewer intervals.

9.5 Gwydir River at Pallamallawa, NSW, Australia, Site number 418001,
1991-01-05 – 1998-07-29



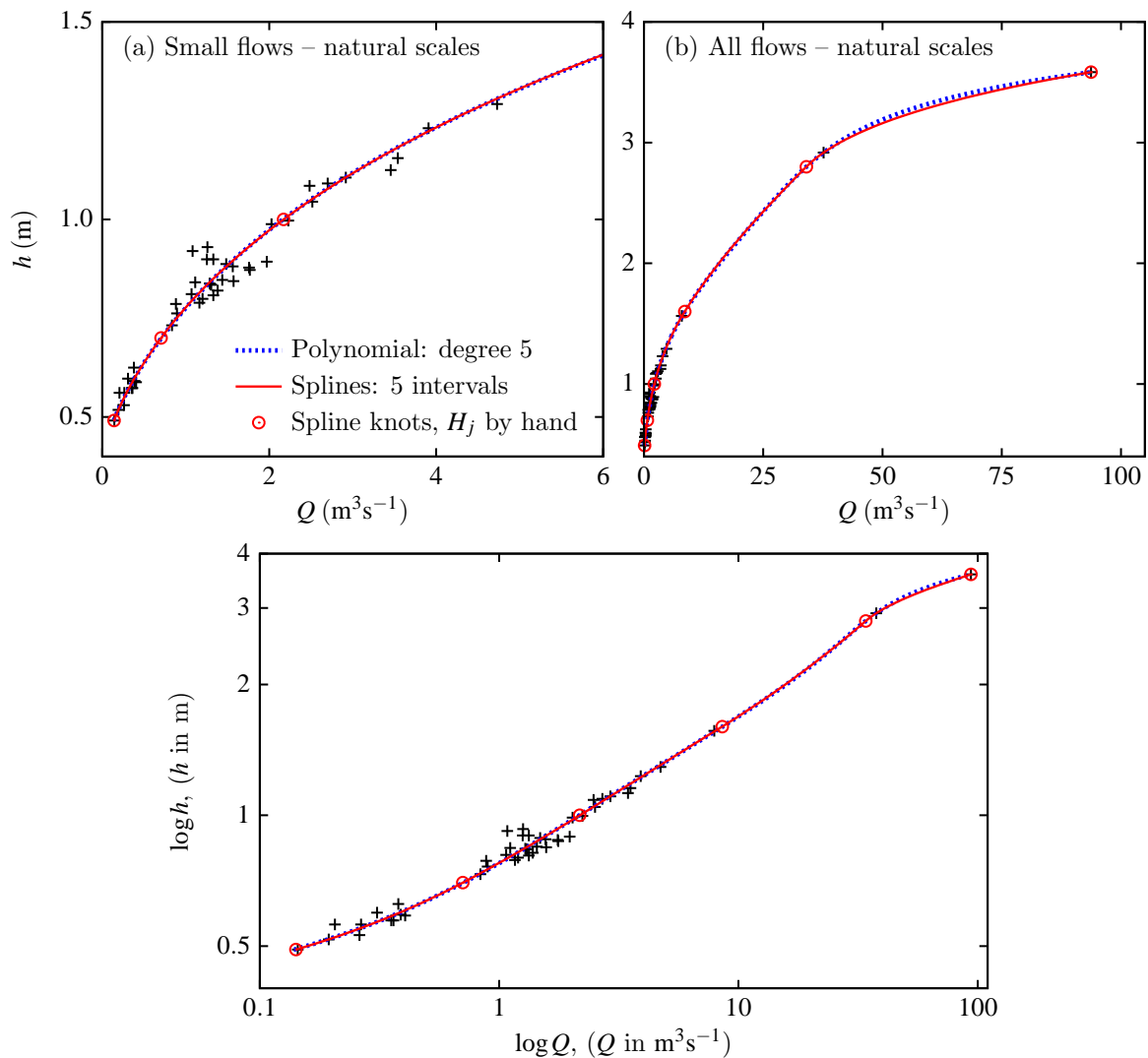
Again, a difficult site from Australia with some large gaps. There are some data points with zero flow, so a plot with a $\log Q$ axis was not possible. Automatic allocation of equally-populated intervals gave good results. A single quadratic from about 2m to 10m meant that the points were approximated, the splines passing smoothly to left and right of the points. The author chose instead, for the first time in the results presented so far, to allocate knots by hand, placing one at 8m, which caused the splines to more-or-less interpolate the points at high flows, whether or not this is justified. Surprisingly, the global polynomial did too. The author has pointed out elsewhere that some of those data points were obtained from a boat floating above flooded agricultural land – maybe a rougher approximation is more justified. In any case one has quite a lot of power to experiment with the approximations.

9.6 Noxubee River near Geiger, AL, USA, USGS Station 02448500, 1970s



This data set has quite a striking ideal form, showing local, channel and overbank control. Both approximations have worked well. When automatic interval allocation was tried with fewer intervals, some irregularities in the data were not picked up (whether they should have been?). That was remedied, first by using as many as 14 intervals, and then, as shown here, by allocating knot locations by hand – remarkably with only five quadratic splines – the whole structure is described, with regions of rapid variation. Also the global polynomial has also worked well here. There are many data points to make it more robust, but the regions of rapid variation must have been a challenge. We seem to have overcome Runge’s phenomenon!

9.7 Wehadkee Creek below Rock Mills, AL, USA, USGS 02339225, 1978-10-06 – 1989-10-17



Both methods work well. This is one case where the automatic allocation of spline knots did not work so well because of large gaps between groups of data points. It was necessary to place a knot or two near isolated points of high stage. Again, surprisingly in view of those gaps, the polynomial method has been very robust.

9.8 Summary of results

In all seven examples, using the data points $(h_i, \sqrt{Q_i})$, both methods have been able to give good results. The global polynomial method used Chebyshev polynomials for the basis functions $f_m(h)$ to guard against poor convergence properties of the series. It only requires the specification of degree M , so it almost runs automatically. What is not automatic is the choice of that M , or more properly, the M at which to stop before unacceptable fluctuations occur. The approximating splines used quadratic approximation $M = 2$ throughout. Automatic allocation of the interior spline knots H_j , $j = 2, \dots, J$ worked completely satisfactorily in four of the seven cases, thereby also making the rating curve generation almost automatic, requiring only the specification of the number of intervals J . For the other three examples, better results could be obtained by allocating the knots by hand in regions of curvature or for isolated data points for high flows. Based on these results, it can be concluded that both methods are quite satisfactory. Both could be routinely implemented together.

10 Rating curve changes with time

Summary: A feature of both the global polynomial method and the approximating splines method is that the importance of each point can be changed by attaching a weight to it. Points can be weighted according to their age, so that most recent points are given larger weights. In this way, the most recent gaugings can be rationally incorporated to give the most recent rating curve. Another way of proceeding is to consider time to be another variable and to develop a bi-dimensional approximation scheme, polynomial or spline in stage and polynomial in time. This identifies the behaviour of all gaugings at a site.

Above we have developed methods to calculate approximations to all the rating data for a particular station. Now we allow for changes in time. We explore two ways of doing this.

10.1 Weighting points according to their age

One simple method is to calculate the weights w_i to be used in the global-polynomial and approximating-spline methods according to some function of time in the form of the age of the measurement T_i , how long ago it was in the past. The simplest would be to ignore all points older than a year, say:

$$w_i = \begin{cases} 1, & \text{if } T_i \leq 1\text{y} \\ 0, & \text{otherwise} \end{cases} .$$

We tried this on the data presented in Figure 16. For high flows, where there were not enough points to define the curve properly, we obtained poor results, which are not shown.

A better method seems to be to use a smoother function, decaying into the past. This would keep all the points to some extent in determining the shape of the curve. A typical example might be to use the exponential weight factor $w_i = \exp(-\alpha T_i)$, where α is a decay constant. If we write $T_{1/2}$ for the "half-life", the age at which the weight decays by a factor of 1/2, then the expression becomes

$$w_i = \left(\frac{1}{2}\right)^{T_i/T_{1/2}} .$$

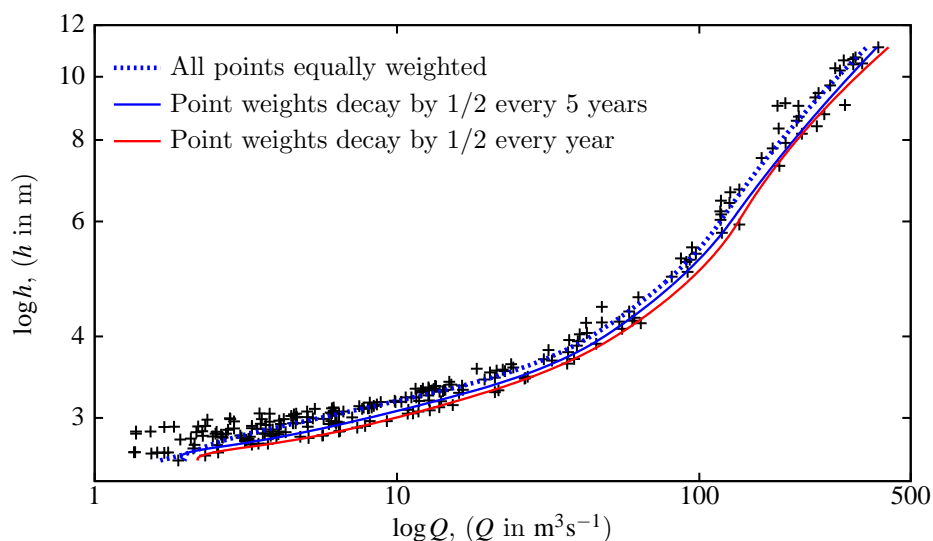


Figure 16: Identifying time variation by using weights that are a function of measurement date – USGS Station 02448500 on the Noxubee River near Geiger, AL, USA from 1984-10-02 to 2015-05-11

We applied the approximating spline method with 6 hand-allocated intervals, using a $\sqrt{Q_i}$ fit, to the data shown in Figure 16. The results seem to bear out our hopes. The differential weighting has shown how the rating curve, and presumably the bed, has moved down over the 31 years, and continues to move. Clearly the latest curve has identified the most recent points and followed them. A global-polynomial approximation, although not plotted, worked just as well. It should be pointed out that the methods have worked rather better at the low flow end than casual inspection might suggest – as always, we are approximating for discharge, namely in a horizontal direction, and the approximating curves pass nicely through the middle of the scattered points horizontally. The log-log scale makes them all look rather more scattered than they really are.

10.2 Using double series in stage and time

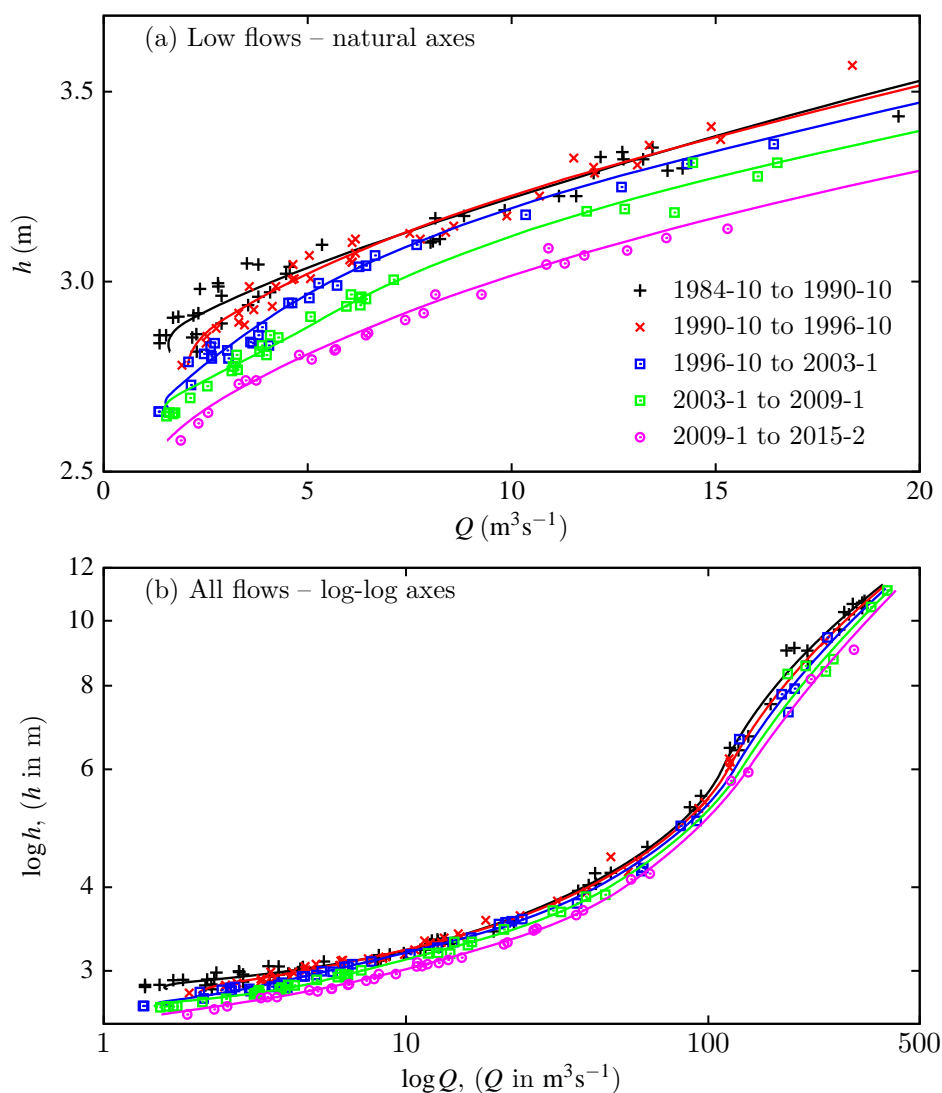


Figure 17: Identifying time variation by using a double series in stage and time – same data as in Figure 16

Usually it is the most recent gaugings that are of interest, and so the previous method seems to work well enough for practical purposes. If one were interested in an examination of how the rating curve varies over time for other purposes, one can use a double series, where the coefficients in the series for global-polynomials and approximating-splines are written as a polynomial in time. Hence, for the global polynomial method one can write the coefficients a_m in equation (15) as $a_m = a_{m0} + a_{m1}t + a_{m2}t^2 + \dots$, with the total sum of the squares e in equation (16) written incorporating this and solved using

optimisation methods. For approximating splines the coefficients $c_{j,m}$ in equation (19) are written as $c_{j,m} = c_{j,m,0} + c_{j,m,1}t + c_{j,m,2}t^2 + \dots$ and the total sum of squares e in equation (22) evaluated, while enforcing modified continuity conditions, equations (21), for each of the corresponding coefficients. We implemented this scheme, and Figure 17 shows results for the same data set as in Figure 16 using a quadratic in time t , in which 31 years of real time was non-dimensionalised to be in the range $[0, 1]$. This method seems to work well also, but it is not so simple as just weighting points according to their age.

11 The rating envelope – a generalisation

***Summary:** A simple model is made of a rating curve under channel control. Effects due to fill and scour of bed material have the (obvious) effect that the whole rating curve is shifted vertically. It is shown that effects of variable resistance can easily dominate such bed level changes, and have the effect of vertically expanding or compressing the curve. Such changes in resistance are ephemeral, as they depend on the arrangement of bed grains and bed forms which can very easily change with flow, and that might be just for a short time. We produce a model explaining the variability of results and show how it is possible to compute upper and lower flow envelopes to the cloud of rating points, thus defining a band within which any combination of flow and resistance, and hence stage, is possible. This is compatible with the idea that any particular flow trajectory, the path of a transient event on rating axes, can form a loop, and the bounds of that loop are the upper and lower bounds of the whole data set. It would be feasible to publish, from routine stage measurements, not only the expected flow from what we think of as the rating curve, but also the largest and smallest flows possible.*

11.1 A model of the effects of channel changes on the rating curve

We develop a model for the effects of changes in cross-section or roughness using a steady uniform channel flow formula, more applicable to channel- and overbank-control in alluvial streams. *Now* we are using hydraulics.

The shape of the rating curve is largely determined by the overall geometry, but fill and scour in the bed will have an effect, as well as changes in the boundary roughness due to changes in the condition of the bed, whether it is armoured, or individual particles exposed and what size they are, as well as changes in vegetation.

It is easier to use the Chézy-Weisbach equation for the details and to provide results for Gauckler-Manning at the end. The equation is

$$Q = \sqrt{\frac{g A^3}{\Lambda P}} S, \quad (24)$$

where we have introduced the symbol Λ as the dimensionless resistance coefficient, which is in fact just a dimensionless representation of Chézy's C , $\Lambda = g/C^2 = \lambda/8$, expressed here as a resistance rather than as a conductance, and where λ is the Darcy-Weisbach resistance coefficient. The other quantities are more familiar: A for cross-sectional area, g for gravitational acceleration, P for wetted perimeter and S for slope.

For modelling purposes we assume a wide stream such that P is equal to the width B , and write $A = B(h - Z)$ where h is the stage and Z is the mean bed elevation:

$$Q = \sqrt{\frac{g}{\Lambda}} B^2 (h - Z)^3 S. \quad (25)$$

In general, the mean bed elevation and the state of the bed will depend on the recent history of the flow. If the flow has been roughly constant at Q for some time, we expect Z to be the equilibrium bed elevation

for that flow and Λ to be the equilibrium resistance coefficient for that flow, so that both are a function of Q .

For our present purposes it seems better to interchange variables and to consider h as a function of Q , so that we obtain results for the effects of channel changes on the quantity that we observe. That is, for a given flow, what are the effects of channel changes on the stage? Hence we re-write equation (25) in terms of h :

$$h = Z(Q) + \left(\frac{\Lambda(Q)Q^2}{gB^2S} \right)^{1/3}. \quad (26)$$

This equation gives us a model rating curve in the form of h as a function of Q . Now, if instead of the equilibrium values of $Z(Q)$ and $\Lambda(Q)$, for that Q , what happens if a flow event has changed the mean bed elevation by ΔZ and the resistance coefficient by $\Delta\Lambda$, while we assume that width B and overall bed slope remain the same? The differential (change) in h is

$$\Delta h = \frac{\partial h}{\partial Z} \Delta Z + \frac{\partial h}{\partial \Lambda} \Delta \Lambda.$$

Substituting the derivatives with respect to Z and Λ obtained by differentiating equation (26) gives

$$\Delta h = \Delta Z + \frac{1}{3} \left(\frac{Q^2}{gB^2S\Lambda^2} \right)^{1/3} \Delta \Lambda. \quad (27)$$

It is simpler if we substitute for the term in brackets using equation (25) to give

$$\Delta h = \Delta Z + \frac{1}{3} (h - Z) \frac{\Delta \Lambda}{\Lambda}. \quad (28)$$

If we use Gauckler-Manning resistance we find a similar result

$$\Delta h = \Delta Z + \frac{3}{5} (h - Z) \frac{\Delta n}{n}. \quad (29)$$

These simple results have some important implications. The first term in each is obvious: if the bed rises by a certain amount, the stage will also rise by that amount – *and is independent of flow so it is the same for all flows*. That is, the rating curve is displaced vertically.

The second term, the change in stage due to resistance changing, is proportional to the fractional change in resistance coefficient, and is multiplied by the mean depth, which shows that change in stage due to change in roughness actually varies with stage, and so the rating curve is stretched vertically. Examining the relative contributions of the terms in equations (28) or (29), we can believe that the changes in bed level are relatively small compared with the overall depth, *i.e.* $\Delta Z \ll h$ and so its effects are small. On the other hand, the relative change of resistance $\Delta\Lambda/\Lambda$ could be finite, and is multiplied by the whole depth, so its effect could be large. Unfortunately we rarely know what the resistance Λ or n is, let alone what change might have occurred in a significant flow event, or how long its effects last.

The author, Fenton (2015b, figure 1.4), has presented a figure which is partly reproduced here as Figure 18 containing results from a number of stream-gaugings from 55 sites in New Zealand. The dimensionless resistance coefficient Λ is plotted against relative roughness (dimensionless grain size) $\varepsilon_{84} = D_{84}/(A/P)$, values for a particular site being joined by thin dotted lines. It can be seen that bed grain size has a large effect on the resistance coefficient. However, the state of the bed does too. The author hypothesised that the approximate lower envelope curve shown red and dashed, corresponded to a co-planar (armoured) bed. The blue dashed curve corresponded roughly to Shields' threshold criterion, above which grains are mobile, but just below which they must be in a random state, with some being fully-exposed. The separation between the two curves shows the effects of relative protrusion or shielding of bed grains. The rather higher values above the blue line show the effects of grain mobility on the resistance – if grains are moving or are suspended, the resistance is understandably greater. The presence or otherwise of bed forms is implicit but unrecognised in the diagram.

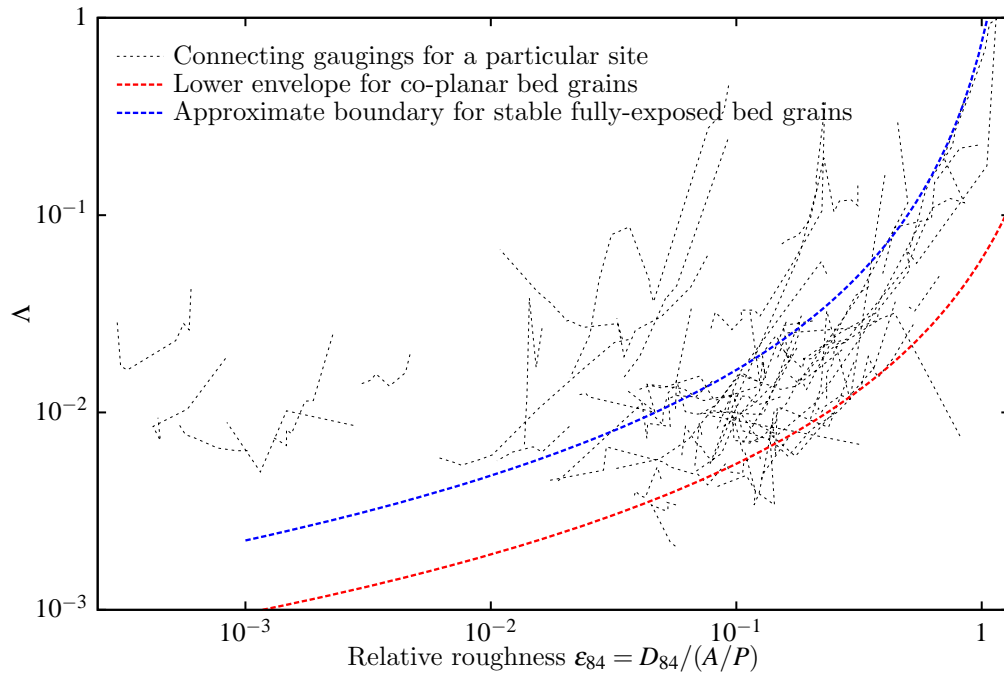


Figure 18: Values of resistance coefficient Λ obtained from gaugings in New Zealand. The thin dotted lines connect the actual data points at a particular site; for clarity, the points themselves are not shown, but their positions are usually obvious.

The most important result for us is to see the vertical extent of data points, how much the resistance Λ varies for a particular grain size, depending on the arrangement of bed grains and possibly whether they are rolling over the bed, or being carried in suspension, not to mention the possibility of bed forms. The resistance of the top dashed curve is about three times that of the lower curve. This shows that the arrangement and stability of the bed is important in determining the resistance. That arrangement is, of course, highly ephemeral. Whereas a flow event in the stream might take some time for fill or scour to take place and to raise or lower the bed, the re-arrangement of surface particles and bed forms, and hence resistance, might take place quickly. Unfortunately we can never really know the state of the bed, for it might continually change with flow and depends on the duration of each flow – the bed can be co-planar and armoured, or individual particles can sit up above the others, or interstices can be filled with smaller particles, or those particles removed, or bed forms can develop, or they can be washed away. We will now consider this in the context of variable flows.

11.2 Unsteadiness and looped rating curves

If there is a rapidly-changing flow event such as a flood, there are methods to deal with the hydraulic effects of flow varying with time (see, for example, §4 of Fenton & Keller 2001). Probably more importantly, however, is the fact that roughness and hence resistance also change relatively quickly, and the relationship between stage and discharge changes with time such that if we were able to measure and plot it throughout the unsteady event we would obtain a looped curve with two discharges for each stage, and *vice versa*, before and after the flow maximum. This is usually described as a *looped rating curve*. The author has usually been sceptical of that term, viewing it more as a *looped flow trajectory on rating axes*.

Let us consider the mechanism by which changes in resistance cause the flood trajectories to be looped, by considering a hypothetical and idealised situation. Such an approach was initiated by Simons &

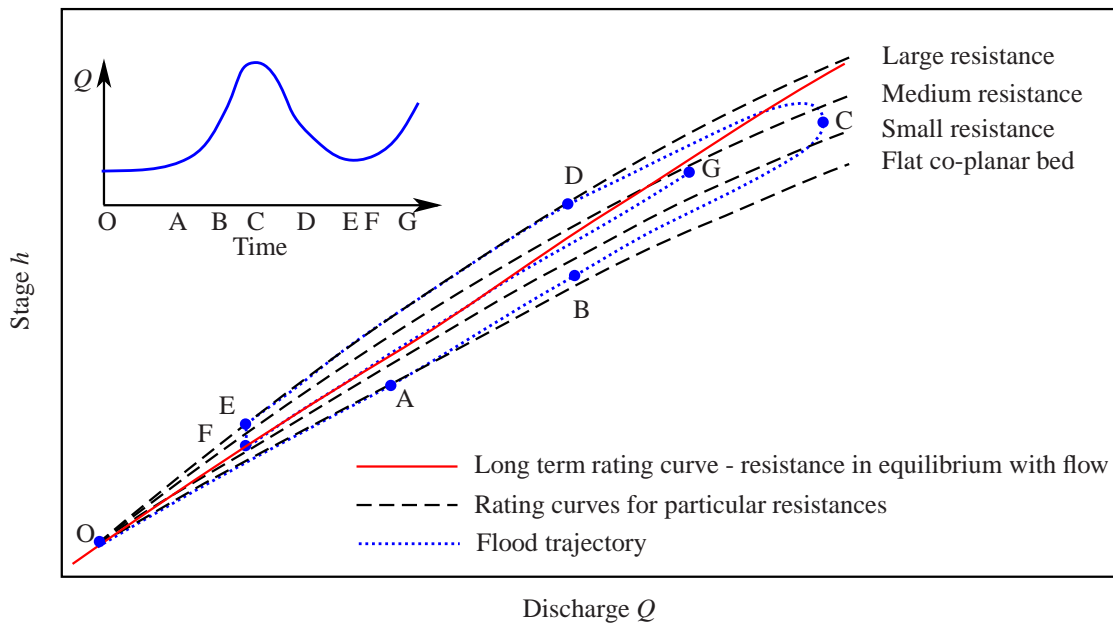


Figure 19: Rating curves for different resistances and a looped flood trajectory showing the possible envelopes to all rating points

Richardson (1962), who were mainly concerned with bed-forms causing the resistance. We do not know how much bed-forms and how much individual grains are responsible for most resistance.

Figure 19 is plotted with rating curve axes, stage versus discharge. The rating curves which would apply if the resistance were a particular value are shown, for a flat bed with co-planar grains, and for various increasing resistances. Also shown on the figure is what have labelled the *long-term rating curve*, on which each point corresponds to the bed resistance in equilibrium with that flow which has lasted long enough such that equilibrium exists.

In the top left corner is a stage-time graph with two flood events, one not yet completed. The points labelled O, A, ..., G are also shown on the *flow trajectory*, showing the actual relationship between stage and discharge at each time.

O: flow is low, over a flat co-planar bed after a period of steady flow.

A: the flow increases. The flow is not enough to change the nature of the bed, and the flood trajectory follows the flat-bed rating curve up to here.

B: the bed is no longer stable, grains move and bed forms develop. Accordingly, the resistance is greater and the stage increases.

C: the flood peak has arrived, and resistance continues to increase, so that a little time later the stage is a maximum.

D: the resistance and bedforms have continued to grow until here, although the flow is decreasing.

E: the flow has decreased much more quickly than the bed can adjust, and the point is close to the instantaneous rating curve corresponding to the greatest resistance.

F: over the intervening time, flow has been small and almost constant, however the time has been enough to reduce the bed-forms and pack the bed grains to some extent. Now another flood starts to arrive, and this time, instead of following the flat-bed curve, it already starts from a finite resistance.

G: after this, the history of the stage will still depend on the history of the flow and the characteristics of the rate of change of the bed.

If we consider a long period of time, with many flow events, we can consider the space between the upper and lower bounds to contain many such flow trajectories, and many individual rating points obtained at moments on those trajectories, leading to a scatter of points between the bounds. The more stable the bed, the less the scatter, and conversely. The longer the flow remains constant before a gauging, the more likely that point will be to fall near the long-term rating curve. By approximating all data points as we have described throughout this work, we actually obtain another rating curve that is not plotted, *the* rating curve as we know it, that which is an approximation to all the data points, and would pass through the region between upper and lower boundaries, possibly not far from the long-term rating curve.

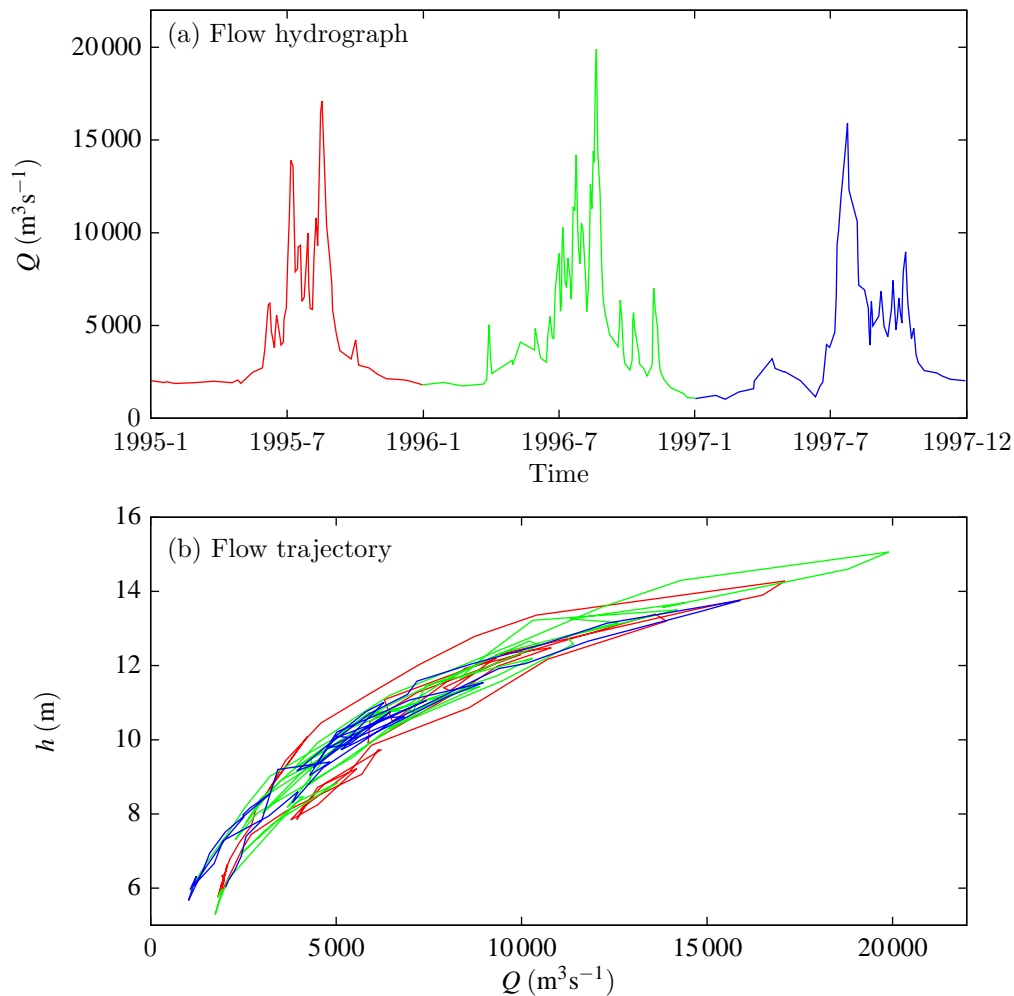


Figure 20: Station 41 on the Red River, Viet Nam, 1995-1997

As an example we have considered the detailed records for 1995 to 1997 for Station 41 on the Red River, Viet Nam, plotted on Figure 20. Part (a) shows the very characteristic flood hydrographs for those years. The flow trajectory plotted on stage/discharge axes in part (b) shows considerable loopiness. Several features resemble those of our idealised version in Figure 19.

11.3 A generalisation – the rating envelope

We suggest an extension of the idea of a single rating curve: that rating data actually falls in a band between a lower boundary, corresponding to smaller resistance and a lower bed, and an upper boundary, where resistance is greater and possibly the bed higher. The more stable the bed, the less the scatter, the narrower the band, and conversely.

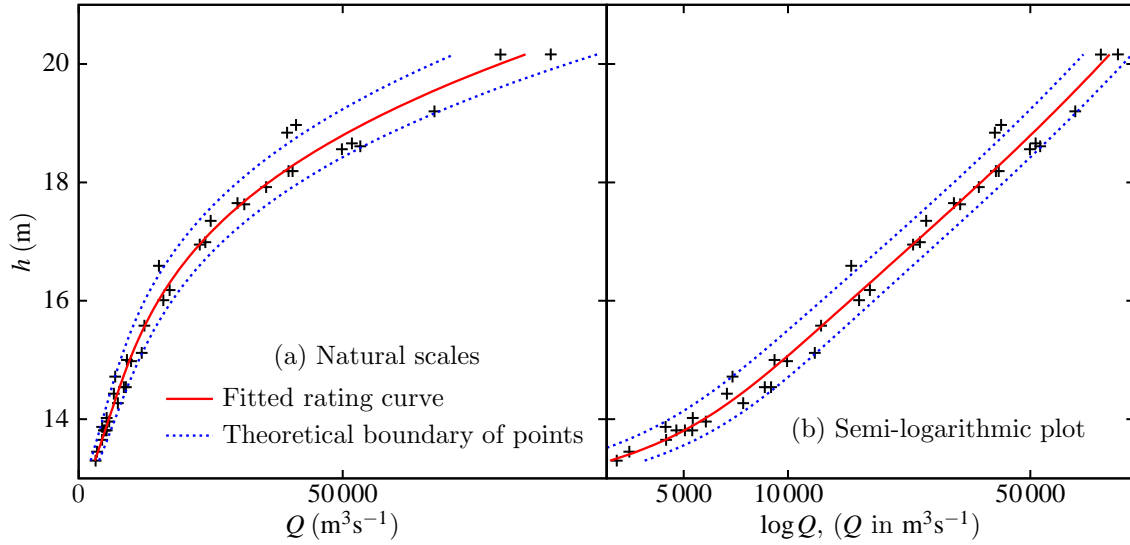


Figure 21: Application of simple model to estimate envelopes to scattered data points, providing some confirmation of the explanation of the scatter. Results for the Brahmaputra River in 1992 from Mirza (2003).

To test how the effects of varying bed level and resistance might affect real gauging results, we considered the results of Mirza (2003) for the Brahmaputra River in 1992, already shown on page 35 but here as shown in Figure 21. We calculated possible envelopes according to our theory above using values of $\Delta Z = \pm 0.1\text{m}$ and $\Delta\Lambda/\Lambda = \pm 0.2$, found by trial and error. They provide quite good bounds to the scattered data points. One might say, neglecting long-term changes with time, that these maximum and minimum envelopes are approximations to the highest and lowest flows likely for each stage

This raises the possibility of the customary plotting of three rating curves and the publication of three different current flows from routine stage measurements: the expected value, the minimum and the maximum likely. Whenever a stage is measured in the future, we can never know what the bed conditions are. The best we can hope to do is use our approximation, plus possibly the variability envelopes, to all or the most recent of the points.

To do this we can think of two methods. We could use the simple method used for Figure 21, obtaining upper and lower envelopes by assuming a constant shift, plus and minus, plus a shift, plus and minus, which would vary linearly with stage. Such results could be computed for any set of data. In keeping with much of our philosophy in this work, however, it would be better to use principles of data approximation to calculate approximations to the upper and lower envelopes of the data points. It is possible to use some of the methods described in this report to calculate them. The author (Fenton 2015a) described how to use approximating splines for that. Global polynomial approximation would work just as well. The method is, first to calculate the approximation to all the points, and to delete those points which lie above or those below the approximation, approximate the remaining points, and repeat as many times as necessary. As approximately half the data points are lost with each pass, the number of passes is limited. The two data points with the smallest and largest values of independent variable h might always be retained to ensure that the final envelope would extend from the smallest to largest values of stage. In practice what one would be doing is approximating the 1/8, say, of all data points which lie furthest from the approximation to all the points. In the spirit of approximation, we can call this an envelope.

If this were done for all sites, there would be a certain uniformity of approach, where the wideness of the band would reflect the tendency towards "loopiness", but instead of that causing despair, we can consider it as providing information as to the range of discharges that can be expected.

Figure 22 shows an example, using the same data as Figure 20 for three years of gaugings at Station 41 on the Red River, Viet Nam. We applied the procedure described above, using values of Q_i itself, and global polynomials with degree $M = 6$. We applied four passes of the halving procedure for each of

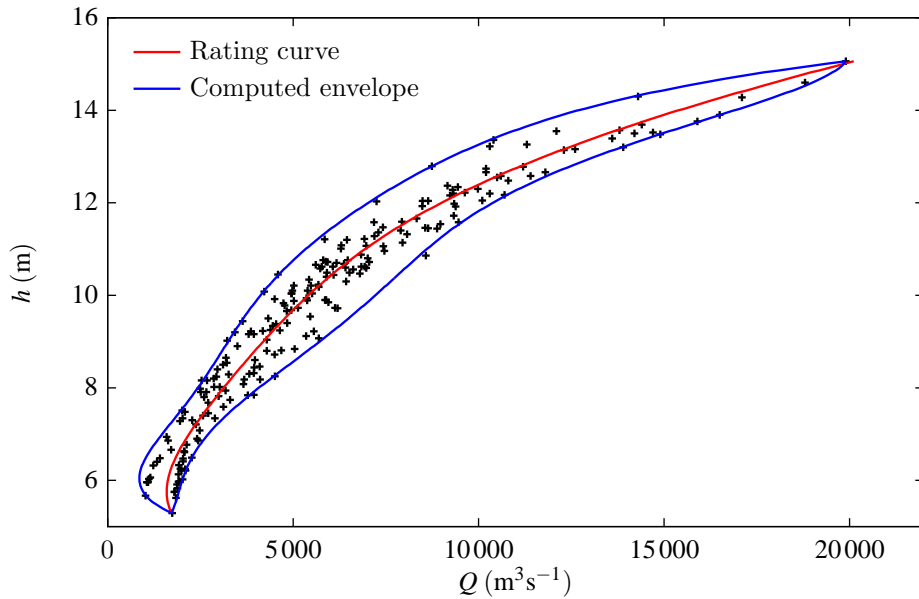


Figure 22: Rating curve and upper and lower bounding envelopes to the data, obtained by the procedure described, using the same data as Figure 20

the upper and lower envelopes, so starting with 217 data points, we ended with about $217/2^4 \approx 15$ for each envelope. It can be seen that the method worked very well. Viewing the points on the plot like this without connecting lines emphasising the loops of the previous figure, the data begins to look rather more like that from any other site. The author believes that it would make some sense to publish routinely the minimum and maximum flows expected for each stage reading.

12 Rating shifts and the incorporation of later measurements

Summary: *The previous section calls into question the current practice of inserting "shift curves", incorporating a new discrepant data point by passing the curve through it but diminishing its effect elsewhere on the rating curve. Our view is that the new point might suggest a redefinition of one of the envelopes of the data, but that might be only short term. The actual rating curve is transitory; the best we can do is to approximate it. Long term changes, on the other hand, can be identified and the curve continually updated using the methods of §10.*

Now we consider the problem of the incorporation of later individual measurements in view of the physical model developed here and the use of global polynomial or spline approximation, possibly with the use of points weighted according to age. There is an extensive discussion of the problem for various types of control in the WMO Manual (WMO 2010), §1.12 *Shifts in the discharge rating*. The advice and procedures are often good and accord with the physical discussion above. There is a general acceptance that a change in bed level will bring about a vertical shift of the rating curve. In some places this, correctly, is interpreted as a vertical shift on a linear axes plot. One can find, however, occasional recommendations for a uniform shift on log-log axes, which is not right. However, the traditional belief in the power function and its imposition of straight lines on log-log axes has limited some of the options.

The extensive discussion of channel control in the Manual includes effects of scour and fill, plus vegetation and bed forms, generalising them as channel "roughness". There is another aspect however, which is not quantified and which might be important, as we have shown above. This is the generalised state of the bed, not just whether bed forms exist or not, but also the arrangement of surface bed grains.

From this we conclude that

- If a measured point is sufficiently distant from the rating curve, then if one believed (a) that that measured point should determine the curve henceforth, and (b) that bed level determines the curve, a first approximation to incorporate the new data point would be to shift the whole rating curve upwards. This would mean that if the new point was δ above/below the curve, one would replace h by $h \mp \delta$ in any of the global polynomial or spline formulae.
- However we suggest that boundary resistance is just as important and quite possibly more important than the actual bed level in determining stage. Ideally then, one should also vertically distort the rating curve to incorporate a change in resistance. Unfortunately that is probably not reasonable, unless we knew two or more recent points that departed consistently from the curve so that we could modify h , not just by a constant δ , but by a linear function of h , as we did for the envelopes in Figure 21.
- In any case, the ephemeral nature of the actual resistance means it is not possible to use such a procedure to predict accurately the stage at any later time. We do not know what flows and bed changes will occur in the future, or will have just occurred up until the moment the curve is required to give us a flow from a routine stage measurement.
- Accordingly we think that there is little need to ascribe major importance to a point not agreeing with the rating curve. Our view of the rating curve is that it is an approximate curve passing through a more-or-less scattered cloud of points, where at least some of that scatter is due to fluctuations in the preceding flows and instantaneous state of the bed when each point was determined. The importance of the rating curve lies in its approximation to a number of points obtained under differing and changing bed conditions which we cannot hope to know.
- This leads us to the sceptical view that any measurement departing from the rating curve does not by itself re-determine the rating curve, as it is affected by the random fluctuations we have described, but also by any steady change in time. It, by itself, is not so important.

One possible way of incorporating a newly-measured point, whether discrepant or not, would be to recompute the rating curve using a large weight w_i solely for that point. Although simpler, that approach would be equivalent to certain existing practices in the profession for incorporating a new data point that is deemed to disagree with the rating curve, where the rating curve is modified to pass through that point, so that it is given the status of uniquely determining the rating curve (which it did on the day of the gauging just for that stage and flow). Current practice then seems to be, rather than to shift the whole curve as one might advocate, to diminish the effect of that point elsewhere on the curve by calculating a new branch of the curve (a "shift curve") which is a straight line on log-log axes between the new point, which it seems to interpolate, and an arbitrarily chosen "hinge point" somewhere further along the rating curve, where the shift curve joins the previous rating curve. The act of interpolation is arbitrary, the choice of hinge point is arbitrary, and of course, the straight line are all arbitrary.

Even if we used the present automatic approximation approach developed in this work with a large w_i for the new point, such a procedure seems incorrect. It distorts the rating curve locally in (Q, h) space to give more weight to that point, tapering its effect off away from the point. Above we have seen that in fact a bed change affects the whole rating curve, even if that might be only for a short period. So, rather than tapering its effect on the rating curve, its effect should be tapered in time.

The best way of incorporating new data points might be to compute the rating curve using weights w_i which are a decreasing function of age of observation, as in §10.1. In this way we recognise that the point's validity might be temporary or it might be part of an overall trend, which the method recognises, and incorporate it as such. If there were enough rating points at a station to define the curve, a relatively short half-life could be used, so as to emphasise the latest data.

13 Conclusions

We have criticised the traditional use of the power function in rating curve approximation. There is no particular reason for any part of the rating curve to follow such a function. Its use has been based on simplified hydraulic formulae. The problem is essentially one of data approximation. The use of the power function has led to unnecessarily labour-intensive methods for the generation of rating curves, with hand manipulation of screen results. A much simpler method is just to use piecewise linear interpolation, whether on natural or logarithmic axes, which we have presented and explained. That requires human intervention to specify a finite number of intervals and at the interpolation points the values of stage and discharge. With the same requirements of human input, cubic splines, also described above, give a more continuous curve.

Technically, a better approach than interpolation is to use approximation by least squares, which does not require the input of visually-approximated discharges at interpolation points. A computationally-robust method for global polynomial approximation has been presented; the nature of the problem requires the use of special polynomials, which are trivially implemented. This technique requires only the input of the degree of polynomial, however monitoring of results for different degrees is necessary, as there is always one degree beyond which the results show unacceptable oscillations.

An even more robust method is the use of splines together with least-squares approximation, for which a methodology has been developed. This requires possibly just the input of the number of intervals, and depending on the results, some manual insertion of computational values of stage at knot points might be necessary, but requires no input of discharge as well.

The approximating methods, global polynomial and splines, both allow the specification of a weight for each data point. These weights can be specified as diminishing with time, such that the method gives what is a better approximation to the most recent data, for which good results have been obtained.

A discussion of the nature of rating data and rating curves has been given, and it has been suggested that the larger-or-smaller scatter of data points might be due to changes in resistance of the stream, as well as local obstructions for low flows. Those changes can be quite ephemeral, so that we never really know what the immediate resistance and hence the rating is. This explains the scatter of the data. A method has been developed to compute upper and lower bounds to the rating data, giving an envelope, so that, for routine stage measurements, not only the most likely discharge, but also the possible maximum and minimum values could be published. This concept seems to bring concepts of loopiness into the framework.

Using interpolation and approximation methods to generate rating curves has been shown to give a family of methods which are simpler, more flexible, and more automatic than those of current practice, opening up some possibilities for more development.

Acknowledgements: I would like to thank Ilaria Prosdocimi and several other people in the water industry for assistance, as well as the United States Geological Survey for the accessibility of its data and openness of its procedures.

References

- Chester, B. L. (1986), Stage-discharge relationships - overview and theory, in *Proceedings, 5th Australian Hydrographic Workshop, Oct. 1986*, Vol. 1, Australian Water Resources Council.
- Conte, S. D. & de Boor, C. (1980), *Elementary Numerical Analysis*, third edn, McGraw-Hill Kogakusha, Tokyo.
- de Boor, C. (2001), *A Practical Guide to Splines*, revised edn, Springer, New York.

- Eilers, P. H. C. & Marx, B. D. (1996), Flexible Smoothing with B-splines and Penalties, *Statistical Science* **11**(2), 89–121.
- Fenton, J. D. (1994), Interpolation and numerical differentiation in civil engineering problems, *Civ. Engng Trans, Inst. Engrns Austral.* **CE36**, 331–337. <http://johndfenton.com/Papers/Fenton94b-Interpolation-and-numerical-differentiation-in-civil-engineering-problems.pdf>
- Fenton, J. D. (2001), Rating curves: Part 2 - Representation and approximation, in *Proc. Conf. on Hydraulics in Civil Engng, 28-30 Nov.*, Inst. Engrns, Aust., Hobart, pp. 319–328. <http://johndfenton.com/Papers/Fenton01Hobart2-Rating-curves-2-Representation-and-approximation.pdf>
- Fenton, J. D. (2015a), Approximating splines and the representation of scattered and not-so-scattered data, Technical report, Alternative Hydraulics Paper 7. <http://johndfenton.com/Papers/Approximating-splines.pdf>
- Fenton, J. D. (2015b), Basic physical processes in rivers, in P. M. Rowiński & A. Radecki-Pawlik, eds, *Rivers – Physical, Fluvial and Environmental Processes*, Springer, chapter 1.
- Fenton, J. D. (2015c), Calculating flow over rectangular sharp-edged weirs, Technical report, Alternative Hydraulics Paper 6. <http://johndfenton.com/Papers/Calculating-flow-over-rectangular-sharp-edged-weirs.pdf>
- Fenton, J. D. & Keller, R. J. (2001), The calculation of streamflow from measurements of stage, Technical Report 01/6, Cooperative Research Centre for Catchment Hydrology, Melbourne. <http://johndfenton.com/Papers/Calculation-of-streamflow-from-measurements-of-stage.pdf>
- Hersch, R. W. (2009), *Streamflow Measurement*, third edn, Taylor & Francis.
- International Standard 7066-2 (1988), *Assessment of uncertainty in the calibration and use of flow measurement devices – Part 2: Non-linear calibration relationships*, International Organization for Standardization, Geneva, also Australian Standard 2360.7.2.
- Kindsvater, C. E. & Carter, R. W. C. (1957), Discharge characteristics of rectangular thin plate weirs, *J. Hydraulics Div. ASCE* **83**(HY6), 1453/1–1453/36.
- Laufer, H. (1996), New findings and recommendations regarding the stage-discharge rating function $Q = K(W - W_0)^E$ with instructions for the use on pocket calculators (In German), *Österreichische Wasser- und Abfallwirtschaft* **48**(3-4), 109–121.
- Mirza, M. M. Q. (2003), The choice of Stage-Discharge relationship for the Ganges and Brahmaputra Rivers in Bangladesh, *Nordic Hydrology* **34**(4), 321–342.
- Morgenschweis, G. (2010), *Hydrometrie – Theorie und Praxis der Durchflussmessung in offenen Gerinnen*, Springer.
- Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992), *Numerical Recipes in C*, second edn, Cambridge.
- Sauer, V. (2002), Standards for the Analysis and Processing of Surface-Water Data and Information Using Electronic Methods, Water-Resources Investigations Report 01-4044, U.S. Geological Survey.
- Simons, D. B. & Richardson, E. V. (1962), The effect of bed roughness on depth-discharge relations in alluvial channels, Water-Supply Paper 1498-E, U.S. Geological Survey.
- WMO (2010), Manual on Stream Gauging Volume II – Computation of Discharge, WMO 1044, World Meteorological Organization, Geneva.
- Wood, S. (2006), *Generalized Additive Models: An Introduction with R*, Chapman and Hall/CRC.