Calibration and Interpolation

John Skilling

Maximum Entropy Data Consultants Ltd., Killaha East, Kenmare, County Kerry, Ireland

Abstract. Interpolation is the problem of fitting a smoothish curve $y(x)$ to data (which may be noisy), and calibration refers to reading $x(y)$ from this curve. This paper presents a fully Bayesian free-form probabilistic solution controlled by the degree of curvature of the interpolant. The optimal interpolant is a cubic spline, accompanied by probabilistic uncertainty and the evidence value.

Keywords: Bayesian, interpolation, calibration, free-form, nonparametric, spline, evidence.
PACS: 02.60.Ed, -2.60.-x, 02.60.Pn

BAYESIAN FORMULATION

We are given measurements (subject to the usual Gaussian noise)

$$y(x_k) = D_k \pm \sigma_k \qquad \text{for } k = 1, 2, \ldots, n \tag{1}$$

at "knots" $x_1 < x_2 < \ldots < x_n$. As analysts, our aim is to infer suitable curves $y(x)$ that can be used to calibrate or interpolate, allowing for uncertainty.

With smoothness in mind, we base our analysis on the curvature prior

$$\Pr(y \mid \phi) \propto \exp(-Q/2\phi)\,, \qquad Q(y) = \int_{-\infty}^{\infty} y''(x)^2\, dx \tag{2}$$

where ϕ is a flexibility parameter to be assigned later. In detail, digitise the x axis into some huge number N of grid points, separated by an arbitrarily small interval. Positions thereby become large integers, which can be scaled back later. Local curvature at grid point i is defined as $y''_i = y_{i+1} - 2y_i + y_{i-1}$ so that the curvature-norm is, in matrix formulation,

$$Q = (\mathbf{y}'')^T(\mathbf{y}'') = \mathbf{y}^T\mathbf{A}\mathbf{y} \tag{3}$$

(**bold** face is used for huge-dimensional vectors and matrices). $\mathbf{A}$ is the 4th-derivative matrix, of arbitrarily high dimension N:

$$\mathbf{A} = \begin{pmatrix} 1 & -2 & 1 & & & & & \\ -2 & 5 & -4 & 1 & & & & \\ 1 & -4 & 6 & \cdot & & & & \\ & 1 & \cdot & \cdot & \cdot & & & \\ & & \cdot & \cdot & \cdot & 1 & & \\ & & & \cdot & 6 & -4 & 1 & \\ & & & 1 & -4 & 5 & -2 & \\ & & & & 1 & -2 & 1 & \end{pmatrix}_N \tag{4}$$

filled with pentadiagonal "1 -4 6 -4 1" rows.

The corners have been arranged so that Q is independent of linear transformations $y \leftarrow y + mx + c$, as a curvature should be. Necessarily, this makes $\mathbf{A}$ singular with two null eigenvalues, so the prior (2) is improper, leading to an unacceptable zero value of the evidence $\Pr(D)$. Eventually, we will evade this difficulty, but for now we suppose some trivial adjustment so that $\det(\mathbf{A})$ will not *actually* be exactly zero. The prior is

$$\Pr(y \mid \phi) = \sqrt{\det(\mathbf{A}/2\pi\phi)}\ \exp(-Q/2\phi) \tag{5}$$

Assuming normal errors, the likelihood for the data is

$$\Pr(D \mid y) = Z^{-1} \exp(-\chi^2/2), \quad Z = \prod_{k=1}^{n} \sqrt{2\pi\sigma_k^2}, \quad \chi^2 = \sum_{k=1}^{n} (y(x_k) - D_k)^2/\sigma_k^2 \tag{6}$$

Vector D and the diagonal correlation matrix σ^{-2} can be used either in n-dimensional context, or in huge-dimensional context with many interleaved zeros. Using the latter, the joint probability underlying our analysis is written as

$$\Pr(\mathbf{y}, \mathbf{D} \mid \phi) = Z^{-1}\ \sqrt{\det(\mathbf{A}/2\pi\phi)}\ \ \exp -\tfrac{1}{2}(\mathbf{y}^T\mathbf{A}\mathbf{y}/\phi + (\mathbf{y}-\mathbf{D})^T\boldsymbol{\sigma}^{-2}(\mathbf{y}-\mathbf{D})) \tag{7}$$

Most probable interpolating function y

We are to minimise Q — or, equivalently, $\mathbf{y}^T\mathbf{A}\mathbf{y}/\phi + (\mathbf{y}-\mathbf{D})^T\boldsymbol{\sigma}^{-2}(\mathbf{y}-\mathbf{D})$ — under constraints at the knots, which we do in stages.

Reduce to finite problem. Constrained variation shows that y'''' is a sum of delta functions at the knots, so y''' is piecewise constant, and y is piecewise cubic with continuous y'', y', y. This alone reduces the initially-huge freedom to a finite list of parameters, which can be written as

$$y(x) = \gamma_0 + \gamma_1 x + \gamma_2 x^2 + \gamma_3 x^3 + \sum_{x_k < x} \lambda_k (x - x_k)^3 \tag{8}$$

where the summation is over knots k to the left of x only. More conveniently, $y(x)$ can be defined by the values $y_k = y(x_k)$ and curvatures $p_k = y''(x_k)$ at the interval edges, these being related through the internal ($k = 2, \ldots, n-1$) second-differences of (8)

$$\frac{y_{k+1} - y_k}{x_{k+1} - x_k} - \frac{y_k - y_{k-1}}{x_k - x_{k-1}} = \frac{x_{k+1} - x_k}{6} p_{k+1} + \frac{x_{k+1} - x_{k-1}}{3} p_k + \frac{x_k - x_{k-1}}{6} p_{k-1} \tag{9}$$

Eliminate curvature. The corners of $\mathbf{A}$ were chosen to annihilate linear functions, so that the optimal interpolant is the "natural" cubic spline being linear beyond the knots (so that $\gamma_2 = \gamma_3 = 0$ in (8) and similarly at the right-hand edge). This linear behaviour implies $p_1 = 0$ on the left and $p_n = 0$ on the right.

In terms of $y_1, \ldots, y_n$, the surviving internal curvatures $p_2, \ldots, p_{n-1}$ are defined by (9), written as

$$Uy = Vp \tag{10}$$

where U is $(n-2)\times n$ tri-diagonal, and V is $(n-2)\times(n-2)$ tri-diagonal operating on the range $2,\ldots,n-1$ only. Because V is non-singular, $p = V^{-1}Uy$. Meanwhile, the curvature is piecewise linear over x, so $Q = \int p(x)^2 dx$ becomes

$$Q = \tfrac{1}{3}\sum_{k=1}^{n-1}(x_{k+1}-x_k)(p_k^2 + p_k p_{k+1} + p_{k+1}^2) = p^T V p = y^T U^T V^{-1} U y \tag{11}$$

involving the same matrix V as before.

Apply the data. Assign optimal $\hat{y}$ by minimising $Q/\phi + \chi^2$, which yields

$$\hat{y} = (U^T V^{-1} U/\phi + \sigma^{-2})^{-1}\sigma^{-2} D \tag{12}$$

This is better written as

$$\hat{y} = D - \sigma^2 U^T (U\sigma^2 U^T + \phi V)^{-1} U D \tag{13}$$

which shows how the optimal points differ from the data in the presence of noise.

This completes the optimal solution: (13) gives $\hat{y}_1,\ldots,\hat{y}_n$, (9) gives $\hat{p}_2,\ldots,\hat{p}_{n-1}$, and $\hat{p}_1 = \hat{p}_n = 0$, then cubic interpolation and linear extrapolation gives $\hat{y}(x)$ anywhere else. Because all the matrices are narrowly banded, all calculation is $O(n)$ fast.

Uncertainty δy

The distribution of $\mathbf{y}$ is governed by the joint probability (7), implying the posterior

$$\Pr(\mathbf{y} \mid \mathbf{D}, \phi) \propto \exp -\tfrac{1}{2}(\mathbf{y}-\hat{\mathbf{y}})^T(\mathbf{A}/\phi + \boldsymbol{\sigma}^{-\mathbf{2}})(\mathbf{y}-\hat{\mathbf{y}}) \tag{14}$$

The uncertainty $\delta\mathbf{y}$ around the optimal interpolant $\hat{\mathbf{y}}$ has covariance

$$< \delta\mathbf{y}\,\delta\mathbf{y}^T > = \; (\mathbf{A}/\phi + \boldsymbol{\sigma}^{-\mathbf{2}})^{-1} \tag{15}$$

From this, we select the point-wise variance

$$< (\delta\mathbf{y}_j)^2 > = \; (\mathbf{A}/\phi + \boldsymbol{\sigma}^{-\mathbf{2}})^{-1}_{jj} \; = \; \phi\frac{\mathrm{adj}(\mathbf{A} + \phi\,\boldsymbol{\sigma}^{-\mathbf{2}})_{jj}}{\det(\mathbf{A} + \phi\,\boldsymbol{\sigma}^{-\mathbf{2}})} \tag{16}$$

of the interpolant at any desired position j.

To evaluate this, begin by calculating the denominator

$$\det(\mathbf{A} + \phi\boldsymbol{\sigma}^{-\mathbf{2}}) = \begin{vmatrix}
1 & -2 & 1 & & & & & & & & \\
-2 & 5 & -4 & 1 & & & & & & & \\
1 & -4 & 6 & -4 & 1 & & & & & & \\
 & 1 & -4 & 6+\alpha & -4 & 1 & & & & & \\
 & & 1 & -4 & 6 & -4 & 1 & & & & \\
 & & & 1 & -4 & 6 & -4 & 1 & & & \\
 & & & & 1 & -4 & 6+\beta & -4 & 1 & & \\
 & & & & & 1 & -4 & 6 & \cdot & \cdot & \\
 & & & & & & 1 & \cdot & \cdot & -4 & 1 \\
 & & & & & & & \cdot & -4 & 5 & -2 \\
 & & & & & & & \cdot & 1 & -2 & 1
\end{vmatrix}_N \tag{17}$$

where $\alpha,\beta,\ldots$ are the individual knot components $\phi\sigma_1^{-2},\phi\sigma_2^{-2},\ldots$ appearing sporadically down the leading diagonal. This will be built up gradually by adding successive rows and columns until the full dimension N is reached.

Determinants that terminate with "1 -4 6 -4 1" corners obey a recurrence relation. Let

$$\Delta_r = \begin{vmatrix} \vdots & \vdots & \vdots & & & & \\ \vdots & P & \vdots & 1 & & & \\ \vdots & \vdots & \vdots & -4 & 1 & & \\ & 1 & -4 & 6 & \cdot & \cdot & \\ & & 1 & \cdot & \cdot & \cdot & 1 \\ & & & & \cdot & 6 & -4 \\ & & & & 1 & -4 & 6 \end{vmatrix}_r \tag{18}$$

be of dimension r. Direct expansion by the last row or column of Δ_r and its associated determinants yields the recurrence relation

$$\Delta_r - 5\Delta_{r-1} + 10\Delta_{r-2} - 10\Delta_{r-3} + 5\Delta_{r-4} - \Delta_{r-5} = 0 \tag{19}$$

for which the general solution is quartic in r, as in the following examples:

$$\begin{vmatrix} 6 & -4 & 1 & & & \\ -4 & 6 & -4 & 1 & & \\ 1 & -4 & 6 & \cdot & \cdot & \\ & 1 & \cdot & \cdot & \cdot & 1 \\ & & \cdot & \cdot & 6 & -4 \\ & & & 1 & -4 & 6 \end{vmatrix}_r = \tfrac{1}{12}(r+1)(r+2)^2(r+3) \tag{20}$$

$$\begin{vmatrix} 1 & -2 & 1 & & & \\ -2 & 5 & -4 & 1 & & \\ 1 & -4 & 6 & \cdot & \cdot & \\ & 1 & \cdot & \cdot & \cdot & 1 \\ & & \cdot & \cdot & 6 & -4 \\ & & & 1 & -4 & 6 \end{vmatrix}_r = 1 \tag{21}$$

The effect of a knot is to change the quartic coefficients from one side to the other: thus

$$\begin{vmatrix} \vdots & \vdots & \vdots & & & & \\ \vdots & P & \vdots & 1 & & & \\ \vdots & \vdots & \vdots & -4 & 1 & & \\ & 1 & -4 & 6+\lambda & -4 & 1 & \\ & & 1 & -4 & \vdots & \vdots & \vdots \\ & & & 1 & \vdots & Q & \vdots \\ & & & & \vdots & \vdots & \vdots \end{vmatrix}_r$$

$$= \begin{vmatrix} \vdots & \vdots & \vdots & & & & \\ \vdots & P & \vdots & 1 & & & \\ \vdots & \vdots & \vdots & -4 & 1 & & \\ & 1 & -4 & 6 & -4 & 1 & \\ & & 1 & -4 & \vdots & \vdots & \vdots \\ & & & 1 & \vdots & Q & \vdots \\ & & & & \vdots & \vdots & \vdots \end{vmatrix}_r + \lambda \begin{vmatrix} \vdots & \vdots & \vdots & & & \\ \vdots & P & \vdots & & & \\ \vdots & \vdots & \vdots & 1 & & \\ & & 1 & \vdots & \vdots & \vdots \\ & & & \vdots & Q & \vdots \\ & & & \vdots & \vdots & \vdots \end{vmatrix}_{r-1} \tag{22}$$

where Q is a "1 -4 6 -4 1" pentadiagonal extension and "λ" at position k is the rightmost knot element contained in Δ_r. Without the extra contribution from the knot, the quartic leftward of k, which can be written as

$$\Delta_s = a_4(s-k)^4 + a_3(s-k)^3 + a_2(s-k)^2 + a_1(s-k) + a_0 \,, \tag{23}$$

would have extended from $s < k$ before the knot straight through to $s > k$ beyond the knot, as in the first determinant on the right. The second determinant on the right gives the increment to the quartic:

$$\lambda \begin{vmatrix} \vdots & \vdots & \vdots & & & \\ \vdots & P & \vdots & & & \\ \vdots & \vdots & \vdots & 1 & & \\ & & 1 & \vdots & \vdots & \vdots \\ & & & \vdots & Q & \vdots \\ & & & \vdots & \vdots & \vdots \end{vmatrix}_{r-1} = \lambda \left(\det P \det Q - \det P_{\_} \det Q^{-} \right) \tag{24}$$

where $P_{\_}$ is P without its last row and column, and Q^{-} is Q without its first row and column. But

$$\begin{aligned} \det P &= \Delta_{k-1} = a_4 - a_3 + a_2 - a_1 + a_0 \\ \det P_{\_} &= \Delta_{k-2} = 16a_4 - 8a_3 + 4a_2 - 2a_1 + a_0 \\ \det Q &= \tfrac{1}{12}(r-k+1)(r-k+2)^2(r-k+3) \\ \det Q^{-} &= \tfrac{1}{12}(r-k)(r-k+1)^2(r-k+2) \end{aligned} \tag{25}$$

Hence, beyond a knot, the quartic form of Δ_r is augmented by the corrective quartic

$$\frac{\lambda}{12}\Big((r-k+1)(r-k+2)^2(r-k+3)\det P - (r-k)(r-k+1)^2(r-k+2)\det P_{\_} \Big) \tag{26}$$

which defines new coefficients a in terms of the old. Thus the partial determinants Δ_r can be stepped past all n knots in only $O(n)$ operations. After the final knot at k^*, Δ_r will take its rightmost quartic form

$$\Delta_r = a_4^*(r-k^*)^4 + a_3^*(r-k^*)^3 + a_2^*(r-k^*)^2 + a_1^*(r-k^*) + a_0^* \tag{27}$$

and direct expansion of more determinants shows that this terminates with the required

$$\det(\mathbf{A} + \phi\boldsymbol{\sigma}^{-\mathbf{2}}) = \begin{vmatrix} \vdots & \vdots & \vdots & & & & & \\ \vdots & \Delta & \vdots & 1 & & & & \\ \vdots & \vdots & \vdots & -4 & 1 & & & \\ & 1 & -4 & 6 & \cdot & \cdot & & \\ & & 1 & \cdot & \cdot & \cdot & 1 & \\ & & & \cdot & \cdot & 6 & -4 & 1 \\ & & & & 1 & -4 & 5 & -2 \\ & & & & & 1 & -2 & 1 \end{vmatrix}_N = 12a_4^* \tag{28}$$

For the adjoint factor in (16), delete the jth row and column of the full matrix to get

$$\mathrm{adj}(\mathbf{A}+\phi\,\boldsymbol{\sigma}^{-2})_{jj} = \begin{vmatrix} \vdots & \vdots & \vdots & & & \\ \vdots & P & \vdots & & & \\ \vdots & \vdots & \vdots & 1 & & \\ & & 1 & \vdots & \vdots & \vdots \\ & & & \vdots & Q & \vdots \\ & & & \vdots & \vdots & \vdots \end{vmatrix}_{N-1} = \det P \det Q - \det P_- \det Q^- \quad (29)$$

where P is the entire matrix before j, including all leftward knots, and Q is the entire matrix after j, including all rightward knots. But

$$\det P = \Delta_{j-1}, \quad \det P_- = \Delta_{j-2} \quad (30)$$

which can be read off from the local quartic which was constructed as part of the calculation of the overall determinant. A similar calculation, starting from the right-hand end and stepping backwards, gives ∇_r, being the partial determinants calculated from the right, giving

$$\det Q = \nabla_{j+1}, \quad \det Q^- = \nabla_{j+2} \quad (31)$$

Thus the variance (*i.e.* uncertainty) of the interpolant can be evaluated at any point by looking up the relevant interval and evaluating a couple of pre-calculated quartic polynomials. (The variance is piecewise 7th order with continuous 3rd derivative.)

Most probable flexibility ϕ

The singularity $\det(\mathbf{A}) = 0$ means that the curvature prior cannot predict all the data. In fact, the prior is — by design — invariant to linear transformation $y \leftarrow y + mx + c$. It *does*, however, predict all the nonlinear structure that lies orthogonal to the subspace of straight lines. Accordingly, we first seek the prior predictive "evidence" *of the nonlinear structure only*, and this will lead us to ϕ. To eliminate the two "$mx+c$" degrees of freedom, we use second differences and consider curvatures p instead of values y.

Likelihood. The n data are given as uncorrelated normal $D \sim \mathrm{N}(y, \sigma^2)$, and we use U from equation (10) to project out the unwanted degrees of freedom. This gives $UD \sim \mathrm{N}(Uy, U\sigma^2U^T)$, written explicitly as

$$\Pr(UD \mid y) = \Pr(UD \mid Uy) = \frac{\exp -\frac{1}{2}(Uy-UD)^T(U\sigma^2U^T)^{-1}(Uy-UD)}{\sqrt{\det(2\pi U\sigma^2U^T)}} \quad (32)$$

Meanwhile, the Jacobian volume-expansion factor in the $(n-2)$-dimensional column space of U is $J = \sqrt{\det(UU^T)}$.

Prior. According to the prior, curvature is assigned as $\mathbf{p} \sim \mathrm{N}(\mathbf{0}, \phi\mathbf{I})$ where $\mathbf{I}$ is the unit matrix, so that $\langle p(x)\rangle = 0$ and $\langle p(x)p(y)\rangle = \phi\,\delta(x-y)$. Values y are the second-integrals

$$y(x) = mx + c + \int_0^x (x-\xi)p(\xi)d\xi \quad (33)$$

Again, second-differencing removes m and c. At the knots, we get

$$\langle (Uy)_k \rangle = 0, \qquad \langle (Uy)_k (Uy)_\ell \rangle = \phi V_{k\ell} \tag{34}$$

where V is the same matrix as in (10). Hence the curvature prior predicts

$$\Pr(Uy \mid \phi) = \frac{\exp(-\frac{1}{2}(Uy)^T V^{-1}(Uy)/\phi)}{\sqrt{\det(2\pi\phi V)}} \tag{35}$$

Selection of ϕ. The joint distribution integrates to

$$\Pr(UD \mid \phi) = \int \Pr(UD \mid Uy) \Pr(Uy \mid \phi)\, d(Uy) = \frac{\exp(-\frac{1}{2}D^T U^T (U\sigma^2 U^T + \phi V)^{-1} UD)}{\sqrt{\det(2\pi(U\sigma^2 U^T + \phi V))}} \tag{36}$$

At small ϕ this tends towards a constant, whereas at large ϕ it is $O(\phi^{-(n-2)/2})$ because the dominating determinant has dimension $n-2$. We suppose a power-law prior for ϕ which accommodates this behaviour, and gives a sensible "best" estimate $\hat{\phi}$. The corresponding ansatz

$$\hat{\phi} = \operatorname*{argmax}_{\phi} \left(\phi^{1/4} \Pr(UD \mid \phi) \right) \tag{37}$$

always gives a sensible result, even when the number of data is only 3 (the minimum needed to detect curvature). The maximum is found by direct one-dimensional search.

Multiple maxima for ϕ

Although each individual factor in the evidence for ϕ changes smoothly and monotonically with ϕ, the combination need not, so that there may be more than one local maximum. When this occurs, it means that some part of the data is favouring one degree of flexibility, while another part is favouring a different value. Usually, this will mean that something has gone wrong with the data, so that the existence of multiple maxima can be used to warn of questionable input values.

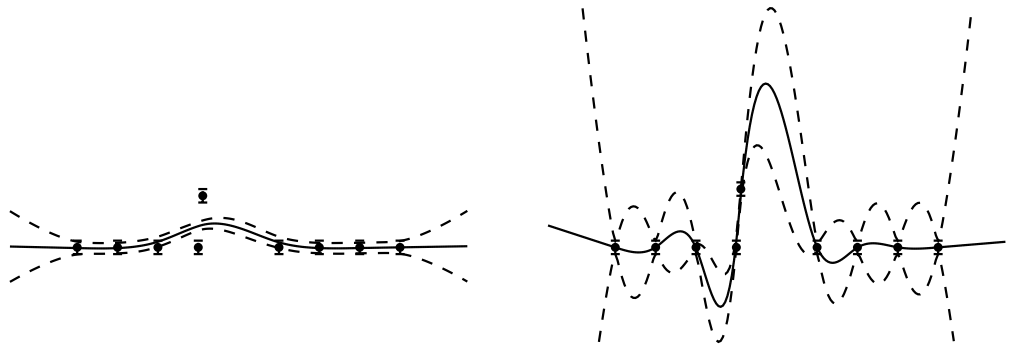

FIGURE 1. Responses (a on left) to weak outlier, and (b on right) to strong outlier. Error bars are $\pm 1\sigma$.

This is illustrated by the algorithm's response to an outlier. Faced with such data, a human operator would have to decide whether or not to accept it. Much the same decision is faced within the algorithm, which detects two locally optimum flexibility values. If the outlier is not too strong, the algorithm stays close to the flexibility it would have had without the outlier, consistently with all the other data, as in Fig. 1a (an optimal interpolant with $\pm 1\sigma$ uncertainties). Eventually, the outlier becomes too strong to ignore, and the results (Fig. 1b) accommodate it by switching to higher flexibility.

Evidence $\Pr(D)$

At the optimal $\hat{\phi}$, the evidence for the differenced data UD is $\Pr(UD \mid \hat{\phi})$. By convention (because $\hat{\phi}$ was not really known in advance), this may be more fairly reported a power of e lower, as $\Pr(UD \mid \hat{\phi})/e$. The Jacobian volume factor then returns this to the subspace "$\perp$" of nonlinearity:

$$\Pr(D_\perp) = \sqrt{\det(UU^T)}\,\Pr(UD \mid \hat{\phi})/e \quad \text{in units of } [y\text{-unit}]^{-(n-2)} \tag{38}$$

To complete the evidence, we need to account for the orthogonal subspace "$\parallel$" of straight lines, which requires the supposedly-prior estimation of two y's along $mx+c$. With linear combinations available of n data with uncertainties σ_k, it is plausible that two useful points would have variances around $(\sum \sigma_k^{-2})^{-1}$. The maximum likelihood of this, attained at optimal m and c, would then be $(1/2\pi)\sum \sigma_k^{-2}$. By convention (because the optima were not really known in advance), this would be more fairly reported 2 powers of e lower, as

$$\Pr(D_\parallel) = (1/2\pi e^2)\sum \sigma_k^{-2} \quad \text{in units of } [y\text{-unit}]^{-2} \tag{39}$$

This treatment is admittedly loose, but it does retain the desired invariance. Finally,

$$\text{Evidence} = \Pr(D) = \Pr(D_\perp)\,\Pr(D_\parallel) \quad \text{in units of } [y\text{-unit}]^{-n} \tag{40}$$

GENERALISATIONS

The curvature prior (2) is a special case of a Gaussian Process (GP), in which the Hessian matrix $\mathbf{A}$ is replaced by (the inverse of) a covariance matrix $\mathbf{C}$, usually with an explicit model $\mathbf{m}$ as well:

$$\Pr(\mathbf{y} \mid \phi, \ldots) = \frac{\exp\left(-(\mathbf{y}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{y}-\mathbf{m})/2\phi\right)}{\sqrt{\det(2\pi\phi\mathbf{C})}} \tag{41}$$

The covariance is usually assigned some standard shape $C(x;x') = F\left((x-x')/w\right)$ where F is normal, Cauchy or other, scaled by a correlation width w. The model, often linear, is inserted because the GP otherwise fails to be invariant to offsets in origin or slope of y. Hence the GP usually requires 4 parameters to be set (ϕ, the linear model, and w).

The *flexibility* ϕ is common to both methods. The GP *correlation width* w has an analogue in the order of smoothness imposed. The curvature prior uses second-order

smoothness based on y'', giving a cubic spline as optimal result. Other orders $y^{(m)}$ could have been chosen, leading to splines of order $2m-1$. However, there is widespread agreement that the cubic spline is "best", lower order being too rough and crude, and higher order being too spread out. Hence we fix the order m at 2, and avoid the extra freedom. If the curvature prior were to be generalised, though, it could be done by altering the curvature-squared "1 -4 6 -4 1" kernel. The *linear model* is explicit in the GP. It is implicit in the curvature approach, being un-necessary for the posterior, but retaining a ghostly presence through its approximated factor in the evidence.

Despite the difficulty of tuning several parameters (and perhaps because generality is thought to be good), a GP prior seems to be the most popular method for probabilistic smoothing. An example is the calibration of radiocarbon dates by Buck *et al.* (2006). These "INTCAL98" data (Stuiver *et al.* 1998) are particularly benign, being quite numerous, equally spaced, and with similar noise throughout.

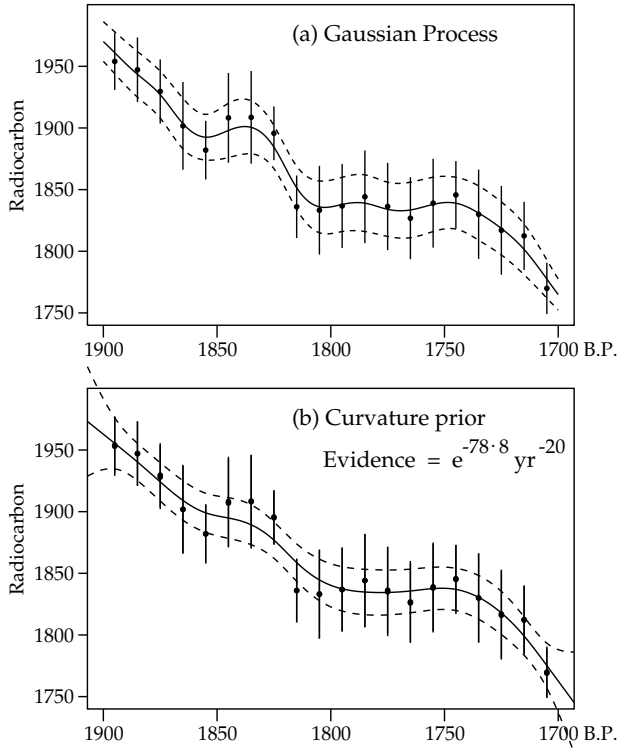

FIGURE 2. Radiocarbon dates *vs* calendar dates Before Present. Results (a) for Gaussian Process from Buck *et al*, and (b) for curvature prior with evidence value. Error bars are drawn at $\pm 2\sigma$.

As might thus be expected, the inferred smooth curves $y(x)$ look fairly similar for the GP (Fig. 2a) and for the curvature prior (Fig. 2b). There is a difference at the edges, because the GP plots the central part of a wider analysis, whereas the curvature calculation used only the 20 visible data-points to illustrate the extra uncertainty of extrapolation outside the data. Other than that, the curvature prior gave a noticeably smoother optimal result, with somewhat smaller uncertainty, and no apparent damage to the data. It costs a factor of 12 in evidence to relax ϕ by the factor of 8 needed to loosen the curvature results to qualitative agreement with the quoted GP.

Unfortunately, it is not possible to give a proper comparison of evidence values, because GP authors do not compute it. Computing the evidence $\Pr(D)$ would admittedly be expensive, because the matrix $\mathbf{C}^{-1}$ is unlikely to be narrowly banded like the curvature Hessian $\mathbf{A}$, so neighbouring data are likely to interfere, making calculation of the huge-dimensional determinant awkward. Instead, GP parameters are tuned by non-Bayesian approximations like cross-validation, and the evidence is ignored.

SUMMARY

Probabilistic calibration and interpolation allow closely-spaced data to merge correctly into a single combination point with average value. This means that extra data will always be of positive help in constraining the results, which is of course exactly what should happen. Such methods should also prove somewhat tolerant of occasional errors in the data, provided these are not too large.

It is no longer necessary to keep calibration points well separated in order to pander to inferior methodology that ignores noise. Neither is it necessary to insist that all calibration signals are strong and hence accurate. It may even be as acceptable to calibrate on a large number of fairly uncertain data as on the traditional small number of unusually clean signals, so the technical requirements of calibration may well simplify.

The results are accompanied at each point by the corresponding uncertainty. This gives an easy check on the completeness of estimation. Excessive uncertainty somewhere warns of inadequate data, so that extra data should be sought in the vicinity. In any case, the curvature-prior algorithm has an in-built diagnostic, warning of incipient trouble from apparently peculiar data.

The curvature prior only needs one parameter, ϕ, which can easily be located. The evidence is available, and shown in the example as a benchmark for other developers. The procedure reduces to classical cubic spline interpolation if its input and output uncertainties are switched off. Relative to this, the extra sophistication costs a factor of 100 or so per data point in the setup phase, and a factor of 3 or 4 per axis point in the run phase. Finally, the whole algorithm is $O(n)$ fast (50 μsec per point on a home PC).

ACKNOWLEDGMENTS

This work was supported by Maximum Entropy Data Consultants Ltd of Cambridge (England), and originally supplied to Micromass Ltd (Manchester UK).

REFERENCES

1. C. E. Buck, D. Gómez Portugal Aguilar, C. D. Litton and A. O'Hagan, "Bayesian nonparametric estimation of the radiocarbon calibration curve", *Bayesian Analysis* **1**, 265–288 (2006).
2. M. Stuiver, P. J. Reimer, E. Bard, J. W. Beck, G. S. Burr, K. A. Hughen, B. Kromer, G. McCormac, L. van der Plicht and M. Spurk, "INTCAL98 radiocarbon age calibration, 24,000-0 cal B.P.", *Radiocarbon* **40**, 1041–1083 (1998).