Gaussian Process Regression for Calibration Curves

Joe Fowler, NIST Boulder Labs

April 14, 2021

Abstract

A calibration curve coverts a microcalorimeter's pulse height measurements into energy estimates. It allows us to generalize from a few energy calibration anchor points to the full range of possible pulse height measurements. We have been using smoothing splines to construct calibration curves. Smoothing splines are piecewise cubic polynomials that avoid exact interpolation of the noisy, measured anchor points in order to reduce curvature in the calibration function. This note shows that cubic smoothing splines are the natural result of a Gaussian Process Regression (GPR) analysis, under certain assumptions about the covariance structure of the Gaussian process. The benefit of taking the GPR point of view is that can estimate the calibration function's uncertainty as well as its expected value, and that it offers a principle for balancing data fidelity against simplicity.

1 Energy calibration for nonlinear microcalorimeters

The response of microcalorimeters to photon energy is not perfectly linear. No matter what statistical summary we use to capture the pulse "size," that quantity invariably grows less than proportionally with the photon energy. This sub-linearity holds for pulse mean, pulse rms value, highest single sample, or optimally filtered pulse height. In this note, I'll use the generic term "pulse height"; it can stand for any of these quantities but normally means the optimally filtered height.

The energy calibration problem requires us to create a function f, where the value $f(p)$ gives the best estimate of the pulse energy when the pulse height is p. We start from a limited set of calibration points (energy-pulse height pairs), along with uncertainties on the values, and estimate a complete $f(p)$ from them. These calibration "anchor points" require a well-known spectral feature to establish the corresponding energy E. This feature might be an atomic fluorescence line, a characteristic decay gamma ray, or an elastic scattering peak observed under monochromatic illumination of a known energy. When the anchor points have non-zero uncertainties in E and/or p, then f need not pass exactly though each one.

1.1 What is splined in Mass before this work

Our Microcalorimeter Analysis Software Suite (Mass) uses cubic splines to construct $f(p)$. However, f itself is not necessarily the cubic spline. For the 2018 metrology analysis [1], we took the pulse height gain $g \equiv p/E$ to be a cubic spline function of pulse height p. The 2013 metrology data [2] was analyzed with $\log E$ as a cubic spline function of $\log p$. Mass supports both methods, and also allows splines to represent either $1/g = E/p$ or $\log g = \log(p/E)$ as functions of p. All four choices have the advantage that finite spline anchor points automatically yield $f(0) = 0$, i.e., zero energy for pulse height of zero.[1]

We chose most recently to spline g as a function of p for the 2018 metrology data, because that choice had the least curvature. That is, assuming g to be exactly linear in p from 5 keV to 7 keV yielded the smallest errors at 6 keV for the low-energy metrology samples, and similarly for 6, 7, and 8 keV with

[1] Technically, this statement would not be true for the $\log E$-$\log p$ spline if it had a negative slope in the region below the energy of any anchor points. A negative slope would imply that E is not a monotone increasing function of p for small pulses, which is unphysical and may be safely ignored.

Method	$h(E,p)$	$f(p)$
Linear	E	p
Loglog	$\log E$	$\log p$
Gain	p/E	p
Inv-gain	E/p	p
Log-gain	$\log(p/E)$	p

Table 1: The five methods currently in MASS for calibration. In each case, the function $h(E,p)$ is approximated as a spline in the function $f(p)$. Function h is readily invertible to solve for E when $h(E,P)$ is known.

the high-energy metrology samples. I argued that is most natural to use the choice that would require the least curvature in the spline. If one of the other spline targets is less curved for your data, MASS lets you use that one. It also supports use of spline for $f(p)$ itself, either with or without the addition of $f(0)=0$ as an extra, implicit anchor point. The five models are listed in Table 1.

A cubic spline with n_k knots has n_k+2 degrees of freedom. What do we do with the extra two? A common choice when there are N values to spline is to interpolate all N points but to designate only $N-2$ of the values as knots; for example, the `scipy.interpolate.InterpolatingSpline` function omits the second and the next-to-last interpolated values from the set of knots. Instead, we use splines with *natural boundary conditions*, which requires the second derivative to be zero at either end of the splined interval. This has the advantage that extrapolation beyond the measured anchor points is relatively benign, as the spline changes only linearly with its argument. The two boundary conditions on the second derivatives consume the two extra degrees of freedom, allowing the number of knots to equal the number of interpolated values, $n_k=N$.

1.2 Approximate calibration in MASS

By default, MASS calibration curves go exactly through each anchor point. This is probably the safer choice for most work. When the utmost care is required in construction of the calibration curve, however, it is important to account for uncertainty in the anchor-point data. We require only that the curve pass "near" the anchor points, near enough to be consistent with their measurement uncertainties. But if the calibration curve need not interpolate the data exactly, what are we to do with the extra freedom?

I studied this problem extensively in 2015 and concluded that the *smoothing spline* was the answer. A smoothing spline does not exactly interpolate the given data; it makes a compromise between fidelity to the data and minimizing the curvature[2] of the spline. The idea is that the more a spline has to curve, the less reasonable a model it is; a very curved spline is probably taking the noise in the data too seriously. A smoothing spline chooses a function h that minimizes the penalty functional ("cost function")

$$C[h] \equiv \sum_{i=1}^{N} \left(\frac{h(x_i)-y_i}{\sigma_i} \right)^2 + \lambda \int_a^b \mathrm{d}x |h''(x)|^2 \tag{1}$$

where x_i and y_i are the N measured anchor point values, σ_i is the measurement uncertainty on y_i, and $h(x)$ is the spline that is meant to approximate $h(x_i)\approx y_i$. We can think of the first term as the usual χ^2 statistic to measure disagreement between data and model, and the integral in the second term is the model's curvature. The curvature integral must cover (at least) the range where f has nonzero curvature, where h is a polynomial of degree 3 instead of 1. Therefore, $a=\min x_i$ and $b=\max x_i$ is usual (though any endpoints that satisfy $a\le x_i\le b$ for all i will give the same result). The regularization parameter λ controls the compromise between fidelity to the data and minimization of curvature. For $\lambda=0$, curvature is not penalized, and minimization of the cost C generates an interpolating spline. In

[2] Here, *curvature* is defined as the integral of the squared second derivative.

the limit $\lambda \to \infty$, curvature is forbidden, and h will be a line: specifically, the line that minimizes the (uncertainty-weighted) sum-of-squared error between data and model.[3]

The use of this smoothing spline for $g(p)$ and calibration curve $f(p) = p/g(p)$ worked well, but there remained two important and unresolved problems. First, how much weight should we give to curvature in the penalty function, i.e., what λ should we use? Second, what is the uncertainty on that calibration curve? Surely it is small near the anchor points and larger for pulses far from any anchor point, but how can we be quantitative?

Until recently, I chose λ by a fairly ad hoc procedure. I reasoned that curvature should be small, but not so small as to make the χ^2 term unreasonably large. Specifically, I chose λ such that the overall penalty $C[h] = N$, which yields a χ^2 value of one per degree of freedom. This choice gives a sensible λ, probably within a factor of two of the "best" value, but I'm sure that we can do better. I hoped that the theory of Gaussian Processes might help with these two problems.

2 Gaussian Processes

A Gaussian Process model supposes that we want to learn what we can about a function $h(x)$ that is a random member of some infinite space of functions. Given some (possibly noisy) estimates y_i of its value at certain points x_i in the domain, we can restrict the plausible set of functions to those that are consistent with the measurements. We can understand the statistical properties of h in a Bayesian framework. Specifically, a Gaussian Process model states that any finite set of samples from $h(x)$ such as $\{h(1), h(5), h(8)\}$ is distributed as a multivariate Gaussian. A Gaussian Process model space is completely specified by its mean and covariance functions, $m(x)$ and $k(x, x')$. Given these, plus the actual measurements[4], we can work out expected values and covariances for any finite set of samples from the unknown function h. Think of $m(x)$ and $k(x, x')$ as the full model space, before measurements; the data allow us to refine these broad functions into a more limited range of expected functions. The refinement of the model given the data is called Gaussian Process Regression (GPR) [3, 4].

The GPR viewpoint on h would immediately solve the second difficulty we have with calibration splines. GPR provides both a model function—the expected value of $h(x)$—and an uncertainty measure—the expected covariance of h at two points (x, x'). Unfortunately, it opens up a much bigger problem than the question of how much to penalize curvature. To get started, we have to choose the mean and covariance functions m and k that specify the GP model space.

Incredibly, *there is a perfectly plausible GP model space that yields smoothing splines as the expected GPR function values.* It is the space of functions that are the sum of a line over the entire real domain plus the integral of a continuous random walk (that is, a once-integrated *Wiener process*) over the interval sampled by the x_i set [5]. Integrating the Wiener process once allows the expected slope to be different at the two ends of the measured interval. Such a GP model has a mean and covariance of

$$m(x) = \beta_0 + \beta_1 x \quad \text{and} \tag{2}$$

$$k(x, x') = \sigma_f^2 [v^3/3 + v^2|x - x'|/2] \tag{3}$$

where $v \equiv \min(x, x')$. The parameters (β_0, β_1) of the mean can be assumed to have a "diffuse" or uninformative Bayesian prior, so that all possible lines are equally probable. The parameter σ_f^2 that scales the covariance function controls the expected amount of curvature—a larger σ_f^2 has higher curvature.

I won't try to argue that a line plus the integral of a Wiener process are *of course* the only plausible Gaussian Process model to apply to the gain vs pulse height function, so smoothing splines are the inevitable result. This statement might be true, but I don't know how to make this case. Let me take a weaker viewpoint instead: that this GP space is reasonable way of modeling functions that have potentially different slopes to the left and right of the measurements, and as little structure as possible

[3] If the data can be exactly interpolated by a line, then that line is found for any value of λ; this case is not encountered in real calibration data.

[4] We also require a model for the noise on estimates of y_i measurements. We will use the simplest possible model: that the noise is independent and Gaussian-distributed with mean zero and variance σ_i^2.

in the measured interval. Given that the space is perfectly plausible *and yields the same spline functions we have already shown to work quite well*, it is the model space we should use for calibration curves.

We have learned that the smoothing spline is a GPR result under a particular model of the Gaussian Process. What overall variance σ_f^2 should we use for the Wiener process (random walk) component in Equation 3? This is equivalent to the question: what curvature penalty λ should be used in the smoothing spline optimization of Equation 1? Here, the Bayesian view of GPR comes to the rescue. Rasmussen & Williams [4] give an expression for the *marginal likelihood*, i.e. the likelihood after integrating (marginalizing) over the values of the spline function at its knots. This marginal likelihood is a function of σ_f^2 alone. It appears as R&W Equation (2.45) and below as Equation 13. The value of σ_f^2 that maximizes the marginal likelihood is the value most consistent with the data. R&W also show that our regularization parameter λ (Equation 1) and the curvature scale σ_f^2 (Equation 3) are related by $\lambda = 1/\sigma_f^2$.

3 Details: the calibration curve as a Gaussian Process

Equation 2 shows that the non-zero mean of the Gaussian Process is linear in the parameters (β_0, β_1). R&W calls this problem *GPR with a basis set.* The results of a GPR with a basis set can be expressed in a closed form. Suppose a regression will be done with n calibration anchor points and two basis functions. The basis set here is the pair $h_0(x) = 1$ and $h_1(x) = x$. Let $\mathbf{H}$ be the $2 \times n$ matrix whose rows are the basis functions at the n samples x_i. Let $\mathbf{K}$ be the function covariance matrix, with $K_{ij} = k(x_i, x_j)$, and let $\mathbf{K}_y$ be the covariance plus measurement noise, so $\mathbf{K}_y = \mathbf{K} + \mathrm{diag}(\sigma_i^2)$. These matrices depend only on the data locations $(\boldsymbol{x})$ and uncertainty σ_i, not on the data values $\boldsymbol{y}$. Compute:

$$\mathbf{A} = \mathbf{H}{\mathbf{K}_y}^{-1}\mathbf{H}^{\mathrm{T}}, \tag{4}$$

$$\mathbf{C} = {\mathbf{K}_y}^{-1}\mathbf{H}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{H}{\mathbf{K}_y}^{-1}, \tag{5}$$

$$\overline{\boldsymbol{\beta}} = \mathbf{A}^{-1}\mathbf{H}{\mathbf{K}_y}^{-1}\boldsymbol{y}. \tag{6}$$

Call the function being modeled $h(x)$. First, suppose that we need its expected value and variance only at a single location $x_\star$. Let vector $\boldsymbol{h}_\star$ be the values of the basis functions at that location, and $\boldsymbol{k}_\star = k(\boldsymbol{x}, x_\star)$ be the function covariance between all sampled points and new location $x_\star$. Then define $\boldsymbol{r}$ as follows, and the expected function value and variance are:

$$\boldsymbol{r} = \boldsymbol{h}_\star - \mathbf{H}{\mathbf{K}_y}^{-1}\boldsymbol{k}_\star \tag{7}$$

$$\overline{g}(x_\star) = {\boldsymbol{k}_\star}^{\mathrm{T}}{\mathbf{K}_y}^{-1}\boldsymbol{y} + \boldsymbol{r}^{\mathrm{T}}\overline{\boldsymbol{\beta}} \tag{8}$$

$$\mathrm{var}(g_\star) = k(x_\star, x_\star) - {\boldsymbol{k}_\star}^{\mathrm{T}}{\mathbf{K}_y}^{-1}\boldsymbol{k}_\star + \boldsymbol{r}^{\mathrm{T}}\mathbf{A}^{-1}\boldsymbol{r}. \tag{9}$$

If we want to find expected function values and their covariance for multiple points $\boldsymbol{x}_\star$ at once, then we can replace the above three equations with their vector/matrix equivalents. Define the matrices $\mathbf{H}_\star$, $\mathbf{K}_\star$, and $\mathbf{K}_{\star\star}$ so that $H_{\star ij} = h_i(x_{\star j})$ has rows equal to the basis functions at all the $\boldsymbol{x}_\star$ points; $K_{\star ij} = k(x_i, x_{\star j})$; and $K_{\star\star ij} = k(x_{\star i}, x_{\star j})$. Then the three equations above generalize to:

$$\mathbf{R} = \mathbf{H}_\star - \mathbf{H}{\mathbf{K}_y}^{-1}\mathbf{K}_\star \tag{10}$$

$$\overline{g}(\boldsymbol{x}_\star) = \mathbf{K}_\star^{\mathrm{T}}{\mathbf{K}_y}^{-1}\boldsymbol{y} + \mathbf{R}^{\mathrm{T}}\overline{\boldsymbol{\beta}} \tag{11}$$

$$\mathrm{cov}(g_\star) = \mathbf{K}_{\star\star} - \mathbf{K}_\star^{\mathrm{T}}{\mathbf{K}_y}^{-1}\mathbf{K}_\star + \mathbf{R}^{\mathrm{T}}\mathbf{A}^{-1}\mathbf{R}. \tag{12}$$

Wahba has shown [5] that the expected function value $\overline{g}$ (equation 8) is a cubic spline of $x_\star$ with knots at each measured location x_i and with $h''(x) = 0$ at the lowest and highest values of x_i. Therefore, we can use a short cut. We do not have to evaluate Equation 8 directly for every value of $x_\star$ of interest. Instead, we can evaluate the function only n times—at the measured x_i locations (or equivalently, evaluate Equation 11 once at $\boldsymbol{x}_\star = \boldsymbol{x}$). The approximating function we seek, $\overline{g}(x)$, is the unique cubic spline with natural boundary conditions that interpolates these n predictions.

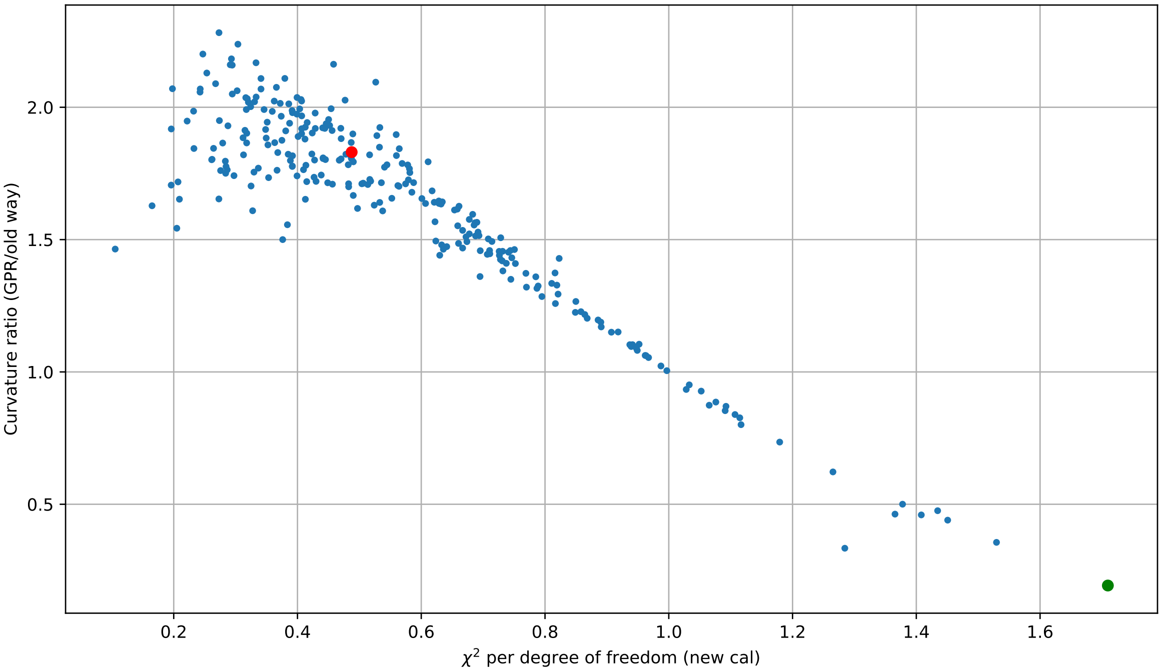

Figure 1: How the new choice of $\lambda = \sigma_f^{-2}$ changes the curvature and the quality of fit. Each dot represents one TES on one of six days of observation from the April-May 2018 metrology experiments (approximately 45–50 TESs were calibrated each day). The majority (93%) of dots are in the upper left quadrant (relative to the lines at 1.0), meaning that the new choice of λ tends to increase the curvature in the cal curve and improve the agreement between data and curve. However, a small fraction of cases went the other way. The two large points correspond to April 25, channel 1 (🔴) and May 1, channel 81 (🟢), whose calibration curves can be seen in Figure 2.

To evaluate the above equations, we need the factor σ_f^2 that scales every evaluation of the function covariance function $k(x, x')$ and thus every entry in any k-matrix: $\mathbf{K}$, $\mathbf{K}_\star$, or $\mathbf{K}_{\star\star}$. We find σ_f^2 by choosing the value that maximizes the marginal likelihood:

$$2 \log p(\boldsymbol{y}|\boldsymbol{x}, \sigma_f^2) = -\boldsymbol{y}^{\mathrm{T}} \mathbf{K}_y{}^{-1} \boldsymbol{y} + \boldsymbol{y}^{\mathrm{T}} \mathbf{C}^{-1} \boldsymbol{y} - \log |\mathbf{K}_y| - \log |\mathbf{A}| - (n - m) \log(2\pi). \tag{13}$$

This probability $p(\boldsymbol{y}|\boldsymbol{x}, \sigma_f^2)$ is the probability of measuring $\boldsymbol{y}$ given the possible β_0, β_1 with a diffuse prior on their values.

4 Comparison of old and new smoothing spline methods

We have been using a (cubic) smoothing spline for g as a function of p in Mass for several years. GPR offers a theoretical basis for this choice, a way to estimate the uncertainty on the curve (Equations 9 or 12), and a marginal likelihood function (Equation 13) that can be maximized to select the smoothing parameter or curvature penalty, $\lambda = \sigma_f^{-2}$.

With this new method to set λ, how much do the calibration curves change? How much does the typical calibrated energy change? Figure 1 shows the April-May 2018 metrology data, with one point per TES per day. By a 12:1 ratio, most calibration curves move in the direction of lower λ, greater integrated curvature, and better agreement between anchor point data and the curve (i.e., lower χ^2). Typical curvature becomes a factor of 1.7 larger, and χ^2 is typically half its previous value of exactly

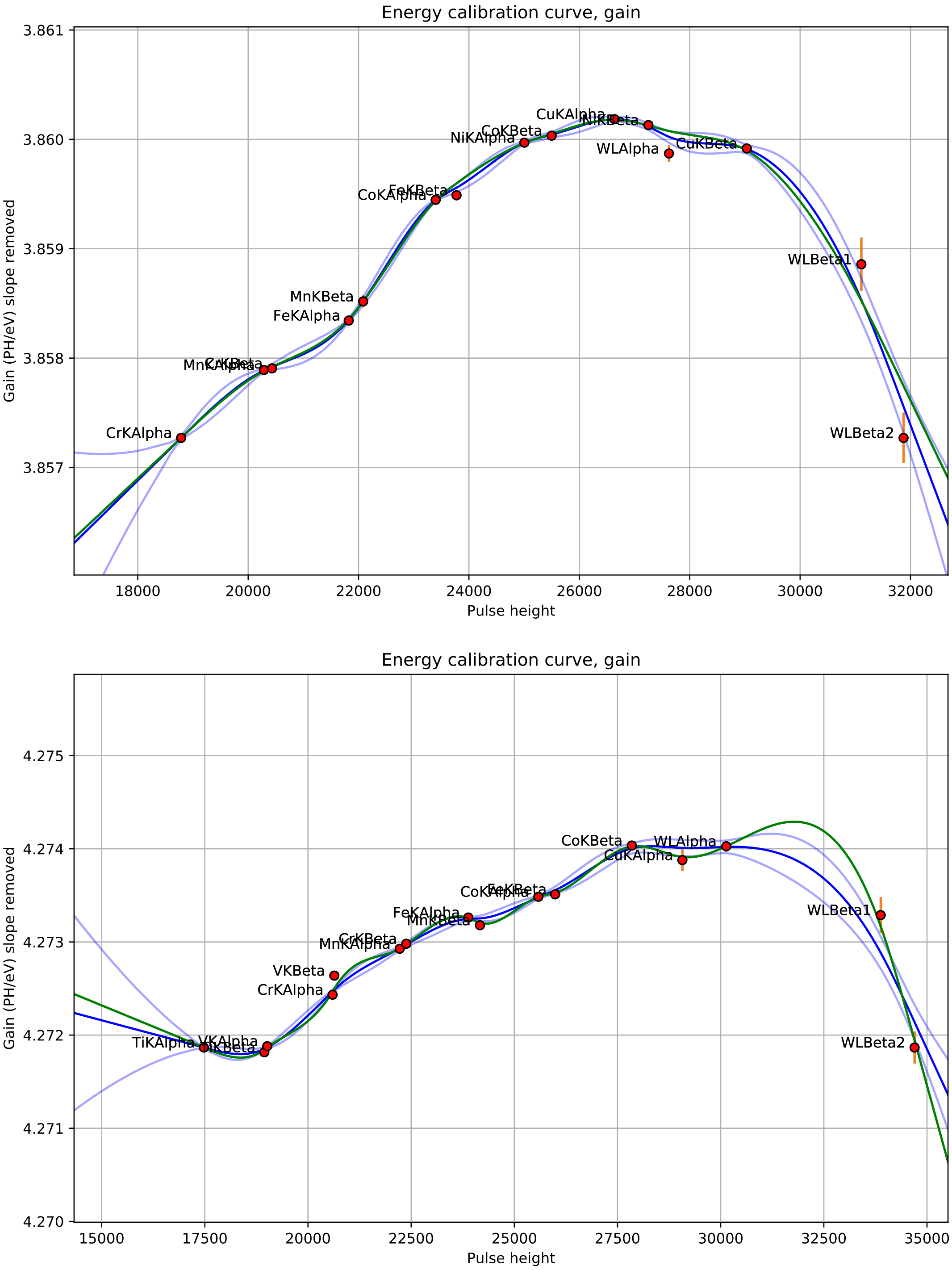

Figure 2: Comparison of old and new calibrations in two cases. Anchor points are the same in both cases. The thick blue curve is the new GPR-based method, with $\pm 1\sigma$ error bands shown in pale blue. The old method in MASS is shown in green. Both examples use the gain method, so $h(E, p) = p/E$ and $f(p) = p$. *Top:* April 25, channel 1. In this instance, like most, the new method results in slightly higher data fidelity and higher cal curvature. *Bottom:* May 1, channel 81. In this instance, the new method results in less data fidelity and lower cal curvature.

1. However, some 7% of calibration curves now favor less curvature and higher χ^2. Figure 2 shows one example of each sort.

The energy shift from the old to the new method is typically less than the systematic uncertainty of 0.15 eV below 8 keV. Figure 3 shows the distribution of energy shifts at each energy. Shifts are large enough to be of concern only in the 9 keV to 10 keV range, where the only anchor points are the low-intensity Lβ lines of tungsten whose absolute energy is not well established.

Figure 4 shows the calibration uncertainties. In general, the GPR estimates of the uncertainties midway between anchor points are larger than our published ad hoc estimate [1], very roughly by a factor of two to four. This is consistent with the new, smaller value of curvature penalty λ. Curvature represents deviations from the simplest, linear $g(p)$ curve, so naturally penalizing it less and having more of it also implies larger uncertainty on the correct calibration model.[5]

Our published calibration uncertainty was estimated by the study of drop-1 and drop-2 cross-validation (CV) tests. We would repeat a calibration without one anchor point—or two, when two were particularly close in energy—and ask how close the new calibration curve was to reproducing the dropped point or points. These results were not as comprehensive as I would have liked, so one explanation for the higher uncertainty in the GPR analysis is simply that the GPR is correct and the ad hoc, published estimate is too optimistic.

Another explanation could be that the number of anchor points in a single TES curve is too small for the marginal likelihood maximization (Equation 13) to work reliably. Perhaps computing λ_{opt} for each TES individually is too susceptible to noise in the limited set of anchor points? If so, we could suppose that a single value of λ or σ_f^2 is appropriate for all TESs, and we could minimize the joint marginal likelihood across all TESs and all days to arrive at a much less noisy result. The 2018 metrology data suggest this approach might work well. The drop-1 CV tests fit with a χ^2 per degree of freedom of 2.3 and drop-2 with 1.2 when λ is estimated individually for each calibration. But when a single λ (the median over all single-curve predictions) is shared, these reduced χ^2 values become 1.5 and 0.7.

In summary, the newly reduced curvature penalty has minimal effect on the calibrated energies in the metrology data, but it suggests that the calibration uncertainty is larger than I had previously thought. Alternatively, it might be that the anchor-point data are too sparse and too noisy for us to optimize the curvature penalty λ separately for each curve. My proposal is to use the per-TES value of λ by default in Mass. We can study the subtle, small trade-offs in the next future high-precision metrology data set.

5 Conclusions

We have considered the energy-calibration curve, specifically the pulse-height gain g as a function of pulse height p, as a Gaussian process. With a plausible assumption about the mean and correlations in that random process, the Gaussian Process Regression approach (Section 2) yields calibration curves equivalent to those we have been using for at least five years: the cubic smoothing spline. The detailed calculation is shown in Section 3. The only update to the smoothing spline procedure is a new method to choose the curvature penalty that any smoothing spline requires.

Comparisons between the old and new choice (Section 4) show that the changes in the calibrated energy are almost always a small fraction of 1 eV at nearly any energy, and are generally less than the systematic uncertainties. However, the estimate of the uncertainty itself is larger than we had previously estimated. I expect that the best way to handle this uncertain uncertainty is to study it carefully when we next make a high-precision metrology measurement. Joint optimization of the curvature penalty across all TESs seems to be a promising approach. For now, the default (per-TES) optimization should supply us very good calibration curves, with the added new feature of automatically estimating the calibration uncertainty at any energy.

[5] Wahba [5] argues that the integrated curvature, the integral in Equation 1, is “a natural measure of the deviation of [the model] from the span of polynomials of degree 1.”

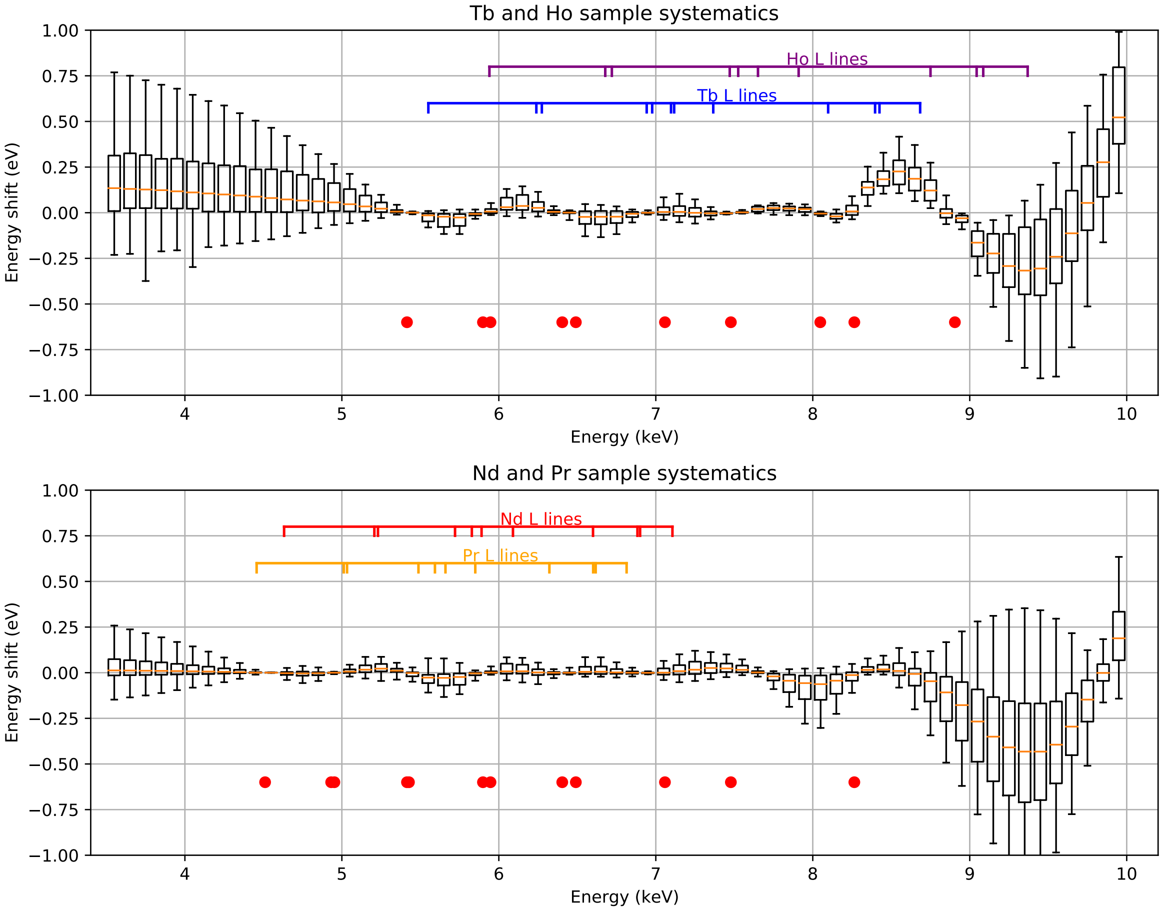

Figure 3: The energy shift (new calibration method minus old) for the 2018 metrology data. The vertical values (whiskers, box, and orange dash) show the {5, 25, 50, 75, 95} percentiles for the shifts in each bin of width 100 eV. *Top:* Ho and Tb samples from April 2018; *Bottom:* Nd and Pr samples from May 2018. Red dots indicate the energy of calibration anchor points. The vertical ticks indicate the several most intense lanthanide L lines. Through the well-calibrated range, the absolute shift is generally less than 0.1 eV. Shifts up to ±1 eV are possible in the 9 keV to 10 keV range, where the anchor points unreliable.

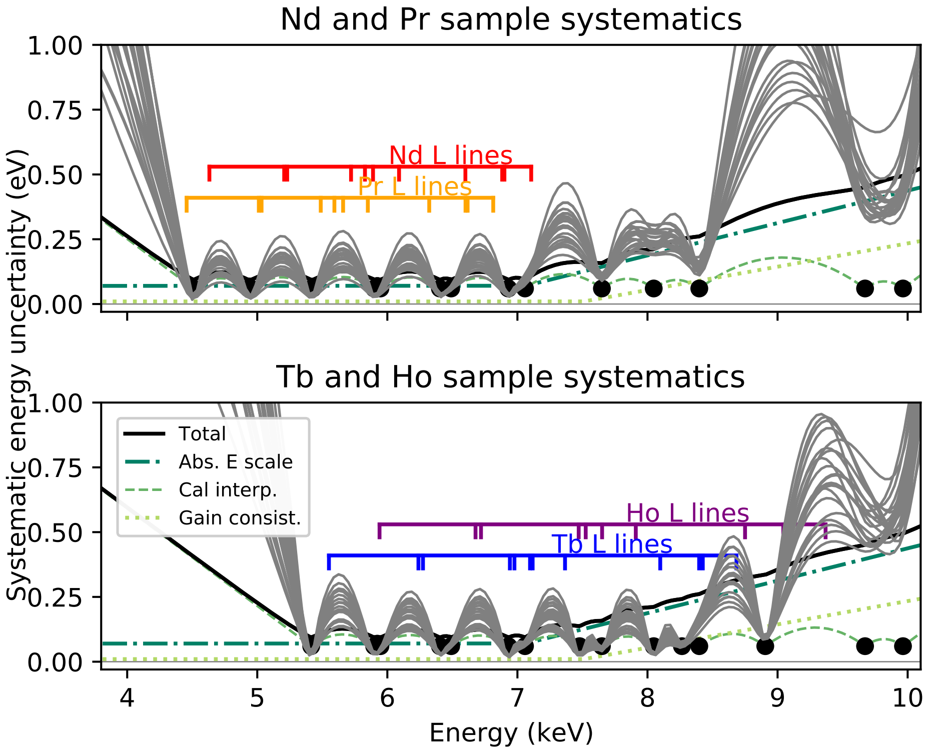

Figure 4: The systematic energy uncertainties on the metrology data (black line), as shown in Figure 17 of the recent metrology paper [1]. Overlaid in gray are the predicted calibration uncertainties from the GPR analysis (the square root of Equation 12) for 20 arbitrary TESs. In general, the GPR uncertainties midway between anchor points are larger than our published ad hoc estimate.

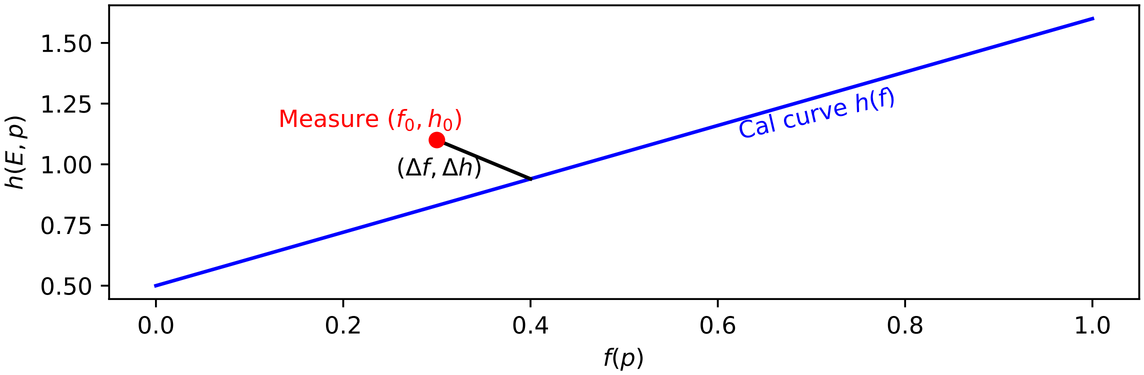

Figure 5: If the cal curve is $h(f)$ and a point (f_0, h_0) not on the curve is measured, what is the appropriate "distance" measure for their separation?

A Correctly combining uncertainties in p and E

Unrelated to the GPR viewpoint on calibration is another calibration issue that I've finally taken the time to get right: the matter of having uncertainties in both the pulse height and energy of the anchor points. The above discussion assumes that the x-values in a GPR (the "locations") are exactly known, and only the y-values have uncertainties. This is a nice simplification, but it's not correct. In all cases, we are trying to *invert* a measurement, so that the x-value is either the pulse height p or its logarithm. There is always an uncertainty on that value. Furthermore, when the y-value being splined is the gain or its inverse or its logarithm, the uncertainties on the x and y coordinates are correlated, because both depend on p. The anchor point energies themselves can also be uncertain. What we need is an *equivalent uncertainty* on the h coordinate, which will be a function of the underlying uncertainties δE and δp.

Here is a leading-order approximation of how these correlated uncertainties work in the calibration curves. I want to work as generally as possible, so let us define the functions h and f such that in whatever calibration model we use, the spline represents $h(E, p)$ as a function of $f(p)$. Naturally, h must be a function readily solved for E given p, if it is to be useful as part of a calibration curve. Also, $f(p)$ must be a monotone function of p. The five possible $\{h, f\}$ pairs now in Mass are named and defined in Table 2.

Suppose a measured anchor point is at (f_0, h_0) in the f-h plane, and the calibration curve is $h(f)$, as in Figure 5. To first order, let that curve be

$$h(f) \approx b + m(f - f_0)$$

for some intercept b and slope m. Consider the possible changes in f and/or h that would bring the measured point to the curve. If the measured data are adjusted by ΔE and Δp, then they would move to $(f_0 + \Delta f, h_0 + \Delta h)$, where $\Delta f = \Delta p\,(\partial f/\partial p)$ and $\Delta h = \Delta p\,(\partial h/\partial p) + \Delta E\,(\partial h/\partial E)$ to first order. Requiring this adjusted point to lie on the first-order approximation to $h(f)$, we can solve for the ΔE compatible with any given Δp:

$$\Delta E = \frac{b - h_0}{\partial h/\partial E} + \frac{m\,\partial f/\partial p - \partial h/\partial p}{\partial h/\partial E}\Delta p.$$

To move the measured value by ΔE and Δp has a "cost" D^2 (square Mahalanobis distance):

$$D^2 = (\Delta E/\delta E)^2 + (\Delta p/\delta p)^2$$

Method	$h(E,p)$	$f(p)$	$\partial h/\partial E$	$\partial h/\partial p$	$\partial f/\partial p$	δh^2
Linear	E	p	1	0	1	$\delta E^2 + m^2 \delta p^2$
Loglog	$\log E$	$\log p$	$1/E$	0	$1/p$	$(\delta E/E)^2 + m^2(\delta p/p)^2$
Gain	p/E	p	$-p/E^2$	$1/E$	1	$g^2(\delta E/E)^2 + (mp-g)^2(\delta p/p)^2$
Inv-gain	E/p	p	$1/p$	$-E/p^2$	1	$(1/g)^2(\delta E/E)^2 + (mp+1/g)^2(\delta p/p)^2$
Log-gain	$\log(p/E)$	p	$-1/E$	$1/p$	1	$(\delta E/E)^2 + (mp-1)^2(\delta p/p)^2$

Table 2: How we compute the combined uncertainty on the y-value of a calibration anchor point, δh, as a function of the uncertainty on the energy δE and pulse height δp. The five models differ in the choice of function $h(E,p)$ being splined, and in the independent coordinates $f(p)$ of that spline. The variables p, E, and $g = p/E$ are the pulse height, energy, and gain at the anchor point. The slope m is the local slope dh/df near the anchor point. The last column gives the equivalent squared uncertainty for each model (Equation 14).

when the E and p uncertainties are δE and δp. If we combine the above two equations to eliminate ΔE, then D^2 is a function only of the p shift, Δp. We can find the specific value of Δp that minimizes D^2, followed by the minimum value of D^2. That value is

$$D^2_{\mathrm{opt}} = \frac{(b-h_0)^2}{(\delta E\, \partial h/\partial E)^2 + \delta p^2 (m\, \partial f/\partial p - \partial h/\partial p)^2}.$$

This tells us that the signal-to-noise distance of the measured point from the calibration curve is the ratio of the y-direction difference to the square root of the denominator above. That is, the square root of the denominator is what we've been seeking, the equivalent error in the h direction:

$$\delta h^2 = (\delta E\, \partial h/\partial E)^2 + \delta p^2 (m\, \partial f/\partial p - \partial h/\partial p)^2. \tag{14}$$

Table 2 gives the relevant functions $h(E,p)$ and $f(p)$, and the resulting δh^2 for each calibration model. These equivalent δh values are now in Mass.

We have a small problem: m is only known after the estimation of the full model, which can only be done after m is already known. In Mass we use an exact (interpolating) spline to estimate m before using Equation 14 to replace that spline with the final, approximating spline. Alternatively, we could iterate a few times between computing the m values and a new approximating spline.

References

[1] Fowler, J. W. et al. "Absolute energies and emission line shapes of the L x-ray transitions of lanthanide metals". *Metrologia* **58** (2021), 015016.

[2] Fowler, J. W. et al. "A reassessment of absolute energies of the x-ray L lines of lanthanide metals". *Metrologia* **54** (2017), 494–511.

[3] Murphy, K. D. *Machine Learning: A Probabilistic Perspective.* MIT Press, 2012.

[4] Rasmussen, C. E. and Williams, K. I. *Gaussian Processes for Machine Learning.* Cambridge: MIT Press, 2006, p. 266.

[5] Wahba, G. "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression". *Journal of the Royal Statistical Society: Series B (Methodological)* **40** (1978), 364–372.