Metafitting: Weight Optimization for Least-Squares Fitting of PTTI Data

R.J. Douglas and J.-S. Boulanger
Time and Frequency Standards Group
Institute for National Measurement Standards
National Research Council of Canada
Ottawa, CANADA K1A 0R6
(613) 993-5186, Fax (613) 952-1394

Abstract

For precise time intercomparisons between a master frequency standard and a slave time scale, we have found it useful to quantitatively compare different fitting strategies by examining the standard uncertainty in time or average frequency. It is particularly useful when designing procedures which use intermittent intercomparisons, with some parameterized fit used to interpolate or extrapolate from the calibrating intercomparisons. We use the term "metafitting" for the choices that are made before a fitting procedure is operationally adopted. We present methods for calculating the standard uncertainty for general, weighted least-squares fits and a method for optimizing these weights for a general noise model suitable for many PTTI applications. We present the results of the metafitting of procedures for the use of a regular schedule of (hypothetical) high-accuracy frequency calibration of a maser time scale. We have identified a cumulative series of improvements that give a significant reduction of the expected standard uncertainty, compared to the simplest procedure of resetting the maser synthesizer after each calibration. The metafitting improvements presented include the optimum choice of weights for the calibration runs, optimized over a period of a week or 10 days.

Introduction

In preparing to fit precision time comparison data, usually questions concerning "optimal" fitting strategies have been addressed in a generic rather than in a specific sense. It is interesting to examine whether, for specific cases, significant advantages might accrue from customizing the fitting strategy to the specific pattern of data points and the noise spectrum. In practice, many really important choices are made before any fit is finalized, and yet are not necessarily optimized as part of the fitting procedure. (a) A fitting metric and method must be chosen (such as least-squares fitting). (b) The set of parameterized basis functions must be chosen: basis function number and type (such as a second order polynomial). (c) An outlier removal method may be adopted (such as iteratively discarding a limited number of points having anomalously high residuals). (d) The relative weighting to be given to each data point must be determined (such as the use of end-point only linear fits vs unweighted linear least-squares fits). (e) A final "consistency of fit with data and noise model" parameter should be derived (such as the reduced χ^2 for a least-squares fit with

PAGE 346 INTENTIONALLY BLANK

known noise). Many of these choices depend in subtle ways on the type of noise encountered, and here precise time measurements often provide details of the noise spectrum which are not trivial to incorporate into an optimal treatment of (a)-(e). Since the term "fitting" is generally interpreted as referring to the determination of a set of parameters from a particular data set, we use the term "metafitting" to encompass the optimization of the broader processes such as (a)-(e).

But in what sense is this "metafitting" to be judged? At first sight, there appear to be too many choices. The fitting might be optimized in an average sense, minimizing some metafitting metric function that sums over experimental residuals. If the autocorrelation function of the noise is known (or modeled) it is possible to calculate and minimize the metric function summing over the expected "residuals" at unmeasured times. Thus the fitting might be optimized in a local sense, minimizing a residual at a specific time (open to choice), or it might be optimized to minimize a residual of the average frequency over a specific interval (each end being open to choice). The point, points or interval must be chosen, and a procedure must be found to estimate the expected residual(s) at times other than those at which measurements have been taken.

Fortunately, international guidelines [1] now strongly suggest a good quantity to optimize: the "standard uncertainty", which is the root-mean-square residual of the fit's extrapolation or interpolation to a specific point, one not necessairly included in the fit. It is also a good quantity to optimize in that a standard method [2](the Wiener-Kolmogoroff theory) exists for any fitting procedure that uses a linear combination of the data. All least-squares fitting procedures with linear coefficients can be handled explicitly in this way [5]. For the purposes of time and frequency metrology, metafitting to minimize the standard uncertainty is a good choice - but it might not be as good a choice in other applications (where, for example it might be more appropriate to try minimizing the occurrence of outlier events having disasterous consequences). For frequency or time interval metrology, the standard uncertainty in the average frequency over an interval makes an even more attractive discriminant for metafitting.

The power law noise models appropriate for PTTI phase comparisons can have low frequency divergences that appear to be worrying to some purists who wish to assure strict stationarity of any process **before** developing its formalism. In the development of computable forms [3],[5], it is straightforward to show that the standard uncertainty in the average frequency of a least-squares fit is not divergent for most commonly encountered power law noise spectra, with the exception of random walk frequency noise. However, the real question is more stringent than simple stationarity: have we enough long-term data on the system being modeled to obtain results which converge? We believe that this type of question can be rigorously handled by imposing a low-frequency cutoff (and thus ensuring a formal stationarity), and then verifying not merely that any results extracted from the model converge as the low frequency limit approaches zero - but also that the results have converged to the desired degree before the low frequency limit is sampling Fourier components of the noise which have not been measured.

Choosing Weights in Weighted Least-Squares Fits

We present here a general strategy for evaluating and optimizing distributions of weights in a weighted least-squares fit to phase data. We will concentrate on optimizing fitting that compares the frequency of a continuously operated oscillator with a frequency standard (perhaps intermittently operated), for the purposes of frequency calibration. The strategy is based upon the analytic expressions for the standard uncertainty in frequency, generally extrapolated over a wider interval than the calibration interval, where a dense set of high-precision phase comparisons would normally be available. A noise model is assumed which has wide applicability to a broad range of frequency standards. The degree of frequency control will be evaluatable for any set of weights in a weighted least-squares fit that is linear in the fitting coefficients (but fully general in the choice of basis functions).

As our metafitting metric we choose the standard uncertainty in the average frequency, evaluated over a general interval which could be considerably displaced from the fitting interval. This is the most appropriate metric for frequency metrology applications, since the standard uncertainty is now the internationally recommended [1] way of specifying calibration uncertainty. With our procedures the standard uncertainty in average frequency can be evaluated for a broad class of noise models, for any set of fitting points, for any extrapolation or interpolation interval, for any linear combination of arbitrary basis functions, and for any least-squares weighting. In particular, in any of the above cases our procedure can evaluate the standard uncertainty the equal-weight procedure (advocated for its robustness) to the end-point procedure (advocated for its "optimum" estimate of frequency for some pure classes of noise), as well as any intermediate case with higher weights near the end-points of the calibration interval. The procedure permits the evaluation of the trade-off of uncertainty for other procedures which are percieved as being more robust. As is shown below, even with large data sets, in some cases it appears to be feasible to choose the optimum set of weights which minimize the standard uncertainty in average frequency for the interval being considered.

Noise Model

The noise model $x_m(t)$ is the modeled phase difference between the master frequency standard and the standard being calibrated. The noise model is taken as being the sum of a deterministic part (which could include a phase offset, frequency offset and frequency drift) and a random noise part, $x_0(t)$. The random noise includes the "full" noise model that is usually used in discussions of frequency standard stability [9]: a sum of five noise processes, each normally distributed about the mean (but with variances which depend on the time sampled in different ways) that have spectral densities of phase noise ($S_x(f)$) that are power laws which range from flat to increasingly divergent at low frequencies. Expressing the five terms in terms of the spectral density of the mean-square of the fluctuations in $\frac{dx_0(t)}{dt}$ (or $y_0(t)$) at a frequency f, $S_y(f)$, each noise term is described by an amplitude h_α which is taken to be independent of any time translations (stationarity and random phase approximations). The sum includes $\alpha = 2$, white phase noise in x; $\alpha = 1$, flicker $(1/f)$ noise in x; $\alpha = 0$, white frequency noise and random walk phase noise; $\alpha = -1$, flicker frequency noise; and $\alpha = -2$, random walk frequency noise. A low-frequency cutoff f_l and an upper frequency cutoff f_h. The spectral density of the mean-square fluctuations in $x_0(t)$ is $S_x(f)$, and for this noise model

$$S_y(f) = \sum_{\alpha=-2}^{2} h_\alpha f^\alpha \quad \text{and} \quad S_x(f) = \sum_{\alpha=-2}^{2} \frac{h_\alpha f^{(\alpha-2)}}{(2\pi)^2} \tag{1}$$

For a given noise model of this type, the standard uncertainty of the fit at any given time can be calculated from the autocorrelation function $\langle x_0(t)x_0(t+\tau)\rangle$. It is divergent for four of our five types of noise unless a low-frequency cutoff is applied, and even then can challenge the accuracy and dynamic range capacities of classical computing. Analytic expressions for this autocorrelation function exist for each type of noise [5], and modern arbitrary-precision computer languages are able to cope directly with the autocorrelation function.

In our analysis of the uncertainty associated with any useful least-squares fit, we expect no divergences to infinity in the standard uncertainty, and so the combinations of the autocorrelation functions must have their divergent parts cancel, with the fitting itself acting as low-frequency cutoff. In considering the standard uncertainty of average frequency from a least-squares fit, we have found it helpful to use analytic expressions [5], [4], [3] for the less divergent general two-interval covariance of the random noise model, that is the covariance of the time-scale departure over the time interval $[t_1, t_2]$ with the time-scale departure over the time interval $[t_3, t_4]$:

$$\begin{aligned} S &= \langle [x_0(t_2) - x_0(t_1)]\,[x_0(t_4) - x_0(t_3)] \rangle \\ &= \int_{t_1}^{t_2} \int_{t_3}^{t_4} \langle y_0(t')\, y_0(t'') \rangle\, dt''\, dt' \\ &= (t_2 - t_1)(t_4 - t_3) \left\langle \bar{y}_{[t_1,t_2]} \bar{y}_{[t_3,t_4]} \right\rangle \end{aligned} \tag{2}$$

where $\left\langle \bar{y}_{[t_1,t_2]} \bar{y}_{[t_3,t_4]} \right\rangle$ is the general covariance of the average frequency: a generalization of the two-sample variance of the average frequency. The generalization includes the possibility of an overlap of the intervals (as well as the possibility of a "dead time" between the intervals), and incorporates the possibility of considering the frequency average over two time intervals of different duration. Just as for the two-sample variance of y, and for the autocorrelation function of $x(t)$, the covariance separates into the five terms of the noise model.

Analytic forms for the five terms of the autocorrelation function of $x(t)$ and for the five terms of the general cross-correlation of $\bar{y}$ are given in references [5], [4] and [3], derived with only the usual assumptions about high and low frequency limits to the noise bandwidth. The references also contain some comments on practical methods for computing values using these forms.

Weighted Least-Squares Fits

Weighted least-squares fitting chooses the n linear coefficients d_l of the n basis functions $g_l(t)$, to arrive at a function $x_p(t)$ which will be used for interpolation or extrapolation. In frequency standards work, we would usually fit a phase offset, a frequency offset, and sometimes a drift rate and higher terms such as daily or seasonal fluctuations.

$$x_p(t) = \vec{d} \cdot \vec{g}(t) = d_1 + d_2 t + d_3 t^2 + \sum_{l=4}^{n} d_l g_l(t). \tag{3}$$

The coefficient vector $\vec{d}$ is chosen to minimize the sum over the N fitting points with phase difference values of $x(t_i)$ at times t_i

$$L_2^2 = \sum_{i=1}^{N} W_i [x(t_i) - \vec{d} \cdot \vec{g}(t_i)]^2 \tag{4}$$

where the weight W_i is applied to the square of the i^{th} residual. Least-squares fitting is done by setting the n derivatives of L_2^2 equal to zero, which gives a set of n linear equations which can be solved for the n fitting coefficients of $\vec{d}$: $\mathbf{G}\vec{d} = \vec{s}$, where $\mathbf{G}$ is an $n \times n$ matrix with elements $G_{qr} = \sum_{i=1}^{N} W_i g_q(t_i) g_r(t_i)$, and $\vec{s}$ is an n-dimensional vector with elements $s_r = \sum_{i=1}^{N} W_i x_0(t_i) g_r(t_i)$. For the purposes of modelling the standard uncertainty, we use $x_0(t_i)$ to model $x(t_i)$, since it can be shown [5] that any general offset in phase, offset in frequency or a linear frequency drift is exactly absorbed by the fit.

Metafitting with Time Uncertainty Metric

One candidate metric for judging weighted least-squares fits is the standard uncertainty in time, determined at a specific time t, relative to the set of fitting points $\{t_i\}$. We can explicitly calculate the effects of the weighted least-squares fit reacting to the noise model for this time t: we are not restricted to studying the variance at the fitting points. The expected variance in $x(t)$ from the fit $\vec{d} \cdot \vec{g}(t)$ can be calculated in terms of the autocorrelation function $left < x_0(t_i) x_0(t_j)$,

$$\left\langle \left[x(t) - \vec{d} \cdot \vec{g}(t) \right]^2 \right\rangle = \sum_{i=0}^{N} \sum_{j=0}^{N} D_i(t) D_j(t) \left\langle x_0(t_i) x_0(t_j) \right\rangle , \tag{5}$$

where $D_0(t) = 1$ and $D_i(t) = W_i \sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i) g_q(t)$. For the standard noise model, the autocorrelation function $\langle x_0(t) x_0(t) \rangle$ can be evaluated analytically [5], although the resulting expressions can challenge the dynamic range of conventional computing. The square root of this variance in $x(t)$ would be the formal metric. The minimization problem, for optimizing this metric with respect to the weights W_i, looks intractable, but for cases of most interest it can be substantially simplified in the same way as is described below for the frequency uncertainty metric.

Variants of this L_2 metafitting metric are also possible, summing variances over multiple test times. Other metafitting metrics of the L_p-norm (Holder norm) class, could also be constructed. The min-max ($\lim p \rightarrow \infty$) norm would minimize the maximum expected time deviation amongst the test times. Metafitting with the $p = 1$ metric would (for this class of metrics) give the most leeway in allowing a small number of test points to have large variances. All these metafitting variants are substantially more intricate to use, and do not readily yield the major computational simplifications which can be found for the single-point L_2 metric.

The method outlined above does not bring any great new insights into optimal ways of combining equivalent clocks, nor for the optimal use of continuously operated primary standards, however when a secondary time scale is to track a primary time scale where only intermittent intercomparisons are available, an optimal choice could be made in terms of the noise processes known to be present.

Metafitting with Frequency Uncertainty Metric

For precise time interval work, where the average frequency is the chief quantity of interest, we wish to minimize the standard uncertainty in average frequency over an interval $[t, t+\tau]$, caused by the noise model as filtered by the weighted least-squares fitting procedure to the points $\{t_i\}$. Although the noise model is independent of time translations, clearly the standard uncertainty in average frequency, u_y, would be expected to depend on the offset of t from $\{t_i\}$, as well as the interval breadth τ. It is defined by

$$u_y^2(t,\tau) \;=\; \left\langle \left[\left\{ \frac{x_0(t+\tau)-x_0(t)}{\tau} \right\} - \left\{ \frac{\vec{d}\cdot(\vec{g}(t+\tau)-\vec{g}(t))}{\tau} \right\} \right]^2 \right\rangle . \tag{6}$$

We note that $\{x_0(t+\tau)-x_0(t)\} = \sum_{j=1}^{N+1}[x_0(t_j)-x_0(t_{j-1})]$, if we define $t_{j=0}=t$ and $t_{j=N+1}=t+\tau$. Although it might be convenient to envisage the set of $\{t_j\}$ as an ordered set with $t_j > t_{j-1}$, it is not necessary to do so. Ordering the fitting points does not detract from the generality in any way, but we do not wish to restrict the values of t or $t+\tau$. We would like to re-express the $\vec{d}\cdot\{g(t+\tau)-g(t)\}$ as a sum over only differences of the form $x_0(t_i)-x_0(t_j)$. We note that we can expand $x_0(t_i) = x_0(t_1) + \sum_{j=2}^{i}\{x_0(t_j)-x_0(t_{j-1})\}$, so that $\vec{d}\cdot\{\vec{g}(t+\tau)-\vec{g}(t)\}$ is equal to

$$\begin{aligned}
&\mathbf{G}^{-1}\sum_{i=1}^{N} W_i x_0(t_i)\cdot\{\vec{g}(t+\tau)-\vec{g}(t)\} = \\
&\sum_{i=1}^{N} W_i \sum_{j=2}^{i}\{x_0(t_j)-x_0(t_{j-1})\}\sum_{q=1}^{n}\sum_{r=1}^{n}(\mathbf{G}^{-1})_{qr}g_r(t_i)\{g_q(t+\tau)-g_q(t)\} \\
&+x_0(t_1)\sum_{i=1}^{N} W_i \sum_{q=1}^{n}\sum_{r=1}^{n}(\mathbf{G}^{-1})_{qr}g_r(t_i)\{g_q(t+\tau)-g_q(t)\}
\end{aligned} \tag{7}$$

and the last term, multiplying $x_0(t_1)$, can be shown to be equal to zero. To show this, it is sufficient to show that $\sum_{i=1}^{N} W_i\vec{g}(t_i)\mathbf{G}^{-1}\vec{g}(t)$ is independent of t, or that $\sum_{i=1}^{N} W_i\vec{g}(t_i)\mathbf{G}^{-1}$ is equal to the vector [1,0,0,...0]. We observe that, from our definition of $\mathbf{G}$ and since $g_1 = 1$, $\mathbf{G}[1,0,0,...0] = \sum_{i=1}^{N} W_i\vec{g}(t_i)$, and premultiplying by $\mathbf{G}^{-1}$ completes this proof, provided only that g_1 is a constant. Thus $u_y^2(t,\tau)\tau^2$ is equal to

$$< [\sum_{j=1}^{N+1} \{x_0(t_j) - x_0(t_{j-1})\}$$

$$- \sum_{i=1}^{N} W_i \sum_{j=2}^{i} \{x_0(t_j) - x_0(t_{j-1})\} \sum_{q=1}^{n}\sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i)\{g_q(t+\tau) - g_q(t)\}]^2 >= \tag{8}$$

$$< [\{x_0(t_{N+1}) - x_0(t_0)\} - \sum_{i=1}^{N} W_i \{x_0(t_i) - x_0(t_{i-1})\} \sum_{q=1}^{n}\sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i)\{g_q(t+\tau) - g_q(t)\}]^2 >$$

Collecting the expressions with the same difference term $\{x_0(t_j) - x_0(t_{j-1})\}$ allows us to write a useful form, namely

$$u_y^2(t,\tau)\tau^2 =< [\sum_{j=1}^{N+1} \tilde{D}_j(t,\tau)\{x_0(t_j) - x_0(t_{j-1})\}]^2 >, \tag{9}$$

where for $2 \le j \le N$, $\tilde{D}_j(t,\tau) = 1 - \sum_{i=j}^{N} W_i \sum_{q=1}^{n} \sum_{r=1}^{n} (\mathbf{G}^{-1})_{qr} g_r(t_i)\{g_q(t+\tau) - g_q(t)\}$; $\tilde{D}_{j=1}(t,\tau) = 1$ and $\tilde{D}_{j=N+1}(t,\tau) = 1$. Multiplying the terms explicitly gives a computable form for the standard uncertainty in average frequency:

$$u_y^2(t,\tau)\tau^2 = \sum_{j=1}^{N+1} \sum_{k=1}^{N+1} \tilde{D}_j(t,\tau)\tilde{D}_k(t,\tau) < [x_0(t_j) - x_0(t_{j-1})][x_0(t_k) - x_0(t_{k-1})] > . \tag{10}$$

The utility of this form lies in the fact that it is a sum over functions of the general form of Eq. 2, which are easier to compute for our full noise model.

Metafitting Weights for Large Data Sets

For a given noise model (defined by the 5 parameters $\{h_\alpha\}$ used to define $S_y(f)$, and a given distribution of fitting points $\{t_j\}$, and for a given interval $[t, t+\tau]$; the standard uncertainty in average frequency over the interval can be calculated: $u_y(t,\tau)$. Thus a choice of weights can be determined which minimizes $u_y(t,\tau)$, the standard uncertainty due to the effects of the random noise. For each fitting point added, another weight must be determined. For small sets of fitting points, the minimization problem is tractable, but for larger sets the minimization appears much less straightforward. The weights could be parameterized to reduce the dimensionality of the problem, at the expense of generality.

The full generality can be retained by largely linearizing the problem. For N fitting points, there are also N weights to choose. Without loss of generality, the set of weights $\{W_i\}$ can be normalized: $\sum_{i=1}^{N} W_i = 1$. If the partial derivative of $\mathbf{G}^{-1}$ with respect to W_k can be constrained to be zero, then most of the N-dimensional search problem can be linearized, leaving a nonlinear search over at worst

$[(n(n+1)/2) - 1]$ dimensions. $\mathbf{G}^{-1}$ will be independent of W_k if each element of $\mathbf{G}$ is constrained to be a constant, $G_{qr} = \sum_{i=1}^{N} W_i g_q(t_i) g_r(t_i)$. Since $G_{qr} = G_{rq}$, and since normally $g_1 = 1$, there remain $[(n(n+1)/2) - 1]$ values. These constraint equations are used in the linear solution, and the optimum values of G_{qr} can subsequently be found by nonlinear searching techniques.

For polynomial fitting, with n basis functions $\{g_k(t) = t^{k-1}\}$, the partial derivative of $\mathbf{G}^{-1}$ with respect to $[N - 2n + 2]$ W_i's there would be only $[2n - 2]$ dimensions for the non-linear search, and if the problem can be set up symmetrically about the time origin, so that the first moment of the weights and all odd moments are zero, there would be only $[n - 1]$ non-linear search parameters. The even moments of the weights (summed over the fitting times $\{t_i\}$) would then be the $[n - 1]$ non-linear search parameters. If the problem is intrinsically asymmetric, then there would be $[2n - 2]$ moments to use as nonlinear search parameters. For extrapolation, it seems clear that there will be little likelihood of driving any W_i negative, but it remains a concern for the general case and must be guarded against.

Consider for example the case of choosing a weighted least-squares fit of a general quadratic to N phase comparison data points at a specific set of times $\{t_i\}$. For a specific noise model described by the coefficients $\{h_\alpha\}$, we want to choose the weights to minimize the standard uncertainty in the average frequency over the time interval $[t, t+\tau]$. By constraining weights to sum to 1, and by constraining the first through fourth moments of the weights to be independent of the first $[N-4]$ weights, we can ensure that $\mathbf{G}^{-1}$ is independent of $[N-4]$ weights. By equating to zero the $[N-4]$ partial derivatives of $u_y^2(t, \tau)$ with respect to W_i we can minimize the standard uncertainty in average frequency with respect to these $[N - 4]$ weights. The easiest form to differentiate for this purpose is one like that of Equation 8, which has collected all the terms multiplied by any weight W_i. Including the constraint equations, we then have N linear equations in the N unknown weights $\{W_i\}$, parameterized in the 4 moments remaining to be searched. The optimized standard uncertainty for this set of four moments is evaluated, and a four-parameter search (each set of moments being optimized by re-solving the N linear equations) this search is tractable by the simplex method (for example). If the problem is symmetric about some time (symmetry for both $\{t_i\}$ and $[t, t+\tau]$), it can be set up so that the first and third moments are zero, and there would be only two parameters to search.

Choosing weights is simpler for a **linear** least-squares fit to N phase comparison data points, taken at a specific set of times $\{t_i\}$. To metafit the best weights that minimize the standard uncertainty in the average frequency over the interval $[t, t+\tau]$ for the noise model of interest, described by the coefficients $\{h_\alpha\}$, we can again linearize the problem - but with only two search parameters (the first and second moments of the weights). We define three constraint equations $\sum_{i=1}^{N} W_i = 1$, $\sum_{i=1}^{N} W_i t_i = M_1$ and $\sum_{i=1}^{N} W_i t_i^2 = M_2$. The N partial derivatives, with respect to the weights, of the standard uncertainty in average frequency over the interval $[t, t+\tau]$ give a set of N equations $\mathbf{F} \cdot \vec{W} = \vec{r}$, where $F_{ij} = (t_i - M_1)\tau/(M_2 - M_1^2) < [x_0(t_i) - x_0(t_1)][x_0(t_j) - x_0(t_1)] >$ and $r_j = < [x_0(t+\tau) - x_0(t)][x_0(t_j) - x_0(t_1)] >$. The first column of $\mathbf{F}$ is a column of zeros. Three of these equations are to be replaced by the three constraint equations: one replacement is for the most ill-conditioned equation j which has t_j closest to the centroid of the weights (M_1) for this iteration, the other two replacements are more arbitrary. If the problem is symmetric about some time (symmetry for both $\{t_i\}$ and $[t, t+\tau]$), it can be set up so that the first moment is zero, and there would be only one parameter to search.

An even simpler case of metafitting is the choice of weights in a simple weighted average, for multiple calibration runs to minimize the standard uncertainty in the average frequency for a specific period, arbitrarily placed with respect to the calibration runs. We consider calibration intervals long enough to be in the regime where the two end point method is chosen for each calibration run, with M such calibration intervals $[t_i, t_i + \tau_i]$. For the weighted average of the M calibrations, the standard uncertainty in the average frequency over an interval $[t, t+\tau]$, $u_y^2(t,\tau)$ is

$$u_y^2(t, t+\tau) \;=\; \Big\langle \Big[\frac{\{x(t+\tau)-x(t)\}}{\tau} - \sum_{i=1}^{M} w_i \frac{\{x(t_i+\tau_i)-x(t_i)\}}{\tau_i}\Big]^2 \Big\rangle. \tag{11}$$

Assigning a weight of -1 to the interval $[t, t+\tau]$, defining τ_0 as being equal to τ, Equation 11 can be rewritten as

$$u_y^2(t, t+\tau) \;=\; \Big\langle \Big[\sum_{i=0}^{M} \frac{w_i}{\tau_i} \{x(t_i+\tau_i)-x(t_i)\}\Big]^2 \Big\rangle. \tag{12}$$

A solution for the optimum weighting procedure is relatively easy to find since the minimum value for $u_y^2(t, t+\tau)$ is to be found for values of w_i satisfying $\frac{\partial}{\partial w_k}[u_y^2(t, t+\tau)] = 0$, so that after taking the derivative and separating out the $i = 0$ term

$$\begin{aligned} \sum_{i=1}^{M} \frac{w_i}{\tau_i\tau_k} \langle [x(t_i+\tau_i)-x(t_i)][x(t_k+\tau_k)-x(t_k)] \rangle \\ = \frac{1}{\tau_k} \langle [x(t+\tau)-x(t)][x(t_k+\tau_k)-x(t_k)] \rangle. \end{aligned} \tag{13}$$

We use $M-1$ of these equations, and for the M^{th} equation we use the normalization equation of the weights: $\sum_{i=1}^{M} w_i = 1$. This gives M simultaneous linear equations in the M unknown weights. The general interval covariance has analytic forms for our noise model, in terms of the $\mathcal{I}$-function [4]. If we define the $M \times M$ matrix $\mathbf{F}$: $F_{1,j} = 1$ for $j = 1..M$, $F_{i,j} = \frac{1}{\tau_i\tau_j}[\mathcal{I}(t_i + \tau_i - t_j)$ $+ \mathcal{I}(t_j + \tau_j - t_i)$ - $\mathcal{I}(t_i + \tau_i - t_j - \tau_j)$ - $\mathcal{I}(t_i - t_j)]$ for $i = 2..M$ and $j = 1..M$, and define $\vec{r}$: $r_1 = 1$ and $r_j = \check{f}rac1\tau\tau_j[\mathcal{I}(t+\tau-t_j) + \mathcal{I}(t_j+\tau_j-t)$ - $\mathcal{I}(t+\tau-t_j-\tau_j)$ - $\mathcal{I}(t-t_j)]$ for $j = 2..M$. The M dimensional weights vector $\vec{w}$ is $\mathbf{F}^{-1} \cdot \vec{r}$.

Applications

For any given potential application of metafitting weights, we must consider whether metafitting is more than an interesting academic exercise: can metafitting find a reduction in the standard uncertainty which is a significant improvement? Since uncertainties are rarely established to better than 10%, an improvement should be larger than this to be deemed significant. Therefore we have examined the simplest case, of linear extrapolation, discussed above, and for the five different power-law noise types we have considered distributions of weights with different moments [6]. We have examined the expected standard uncertainty for both symmetric extrapolation suited to time-scale calibration (where post-processing can be used to apply calibrations from the "future") and to time-assymetric extrapolation suited to real-time applications. For symmetric extrapolation intervals that are large compared to the calibration run's duration, different common weight distributions gave similar uncertainties (differing by less than 10%) except for white phase noise. For one-way extrapolation for times much longer than the calibration run's duration, the uncertainties are even more similar (less than 2% advantage for end-point fitting over equal weights, except for white phase and flicker phase noise). Thus for many PTTI applications, end-point fitting and equal-weight fitting give similar standard uncertainties, and the choice should be between the greater simplicity of the end-point fit and the greater robustness of the equal-weights fitting procedure.

In real-life PTTI work, robustness would often prevail over simplicity. For trying to optimize results from multiple calibration runs, simplicity is valuable to us while robustness is not needed in the model. The optimum processing of a number of calibration runs is expected to be largely independent of the processing within the run.

The main application which has attracted our attention is the optimal use of hydrogen masers, calibrated periodically in frequency with intermittently operated cesium fountain frequency standards [8], [6]. We consider two types of maser operation: free-running and autotuned. We use two power law models for the maser noise, representing a free-running hydrogen maser (type 1) with $h_2 = 2.7\times10^{-24}, h_1 = 2.9\times10^{-30}, h_0 = 2.9\times10^{-27}, h_{-1} = 2.6\times10^{-31}$ and $h_{-2} = 7.2\times10^{-36}$; and an auto-tuned maser (type 2) with $h_2 = 6.7\times10^{-23}, h_1 = 2.9\times10^{-30}, h_0 = 2.9\times10^{-27}, h_{-1} = 7.2\times10^{-31}$ and $h_{-2} = 4.9 \times 10^{-37}$. NRC has two low-flux masers which would benefit from a metafitting optimization of the weights *within* a calibration run of an hour, since there is still some white phase noise contribution for this calibration interval. Preliminary analysis suggests that the end-point procedure is within 10% of the optimum. For phase data taken every 30 s for an hour, extrapolated to an interval of a day, the end point method is 1.2% better than the equal-weight linear least squares fit for our free-running maser model, and as good for the type 2 maser model. Thus we can use the simple two end points procedure to establish the best frequency transfer accuracy for multiple calibration runs. For this procedure the standard uncertainty for multiple calibration runs can be calculated more easily than in the general case.

Within the context of end-point fitting from each calibration run there are still metafitting choices to be made about the way in which the runs are to be used. One possible strategy is a loose lock in frequency: after a calibration run (an hour in duration, in our example) is complete, the frequency of the maser is reset (through the synthesizer control, for example), either immediately - or after some delay. Clearly the least delay is best, and we chose this procedure with zero delay as the reference procedure as we examine a series of possible improvements.

A slightly better possibility might be to have an output tightly locked in phase to the cesium fountain during the calibration run, followed by a frequency lock to the fitted frequency of the calibration run. The phase-lock type of frequency control removes the noise of the maser during the calibration run, giving it an advantage that remains noticeable for extrapolation intervals many times longer than the calibration interval. However, for extrapolations of an hour-long calibration out to a period of a day or more, there is not a large advantage: 2.3% for the free-running maser and 2.4% for the autotuned maser model.

A more significant advantage comes from allowing postprocessing, as can often be tolerated in time-scale construction and for frequency intercomparisons. We consider a single calibration interval t_c and calculate the ratio of the standard uncertainty of the average frequency over an interval τ for the best real-time frequency control to the symmetrically extrapolated time interval τ. The quantitative postprocessing advantage will depend upon the specific processing scheme or schemes envisaged - the duration and frequency of calibration intervals. The postprocessing advantage is up to a factor of two [6].

A postprocessing advantage of two is really quite significant. To achieve the same improvement in the maser ensemble could be done - by increasing the maser ensemble size by four times. The postprocessing advantage of greatest interest to us is for τ representing extrapolation to the time interval between calibrations - which we expect would be between 1 day and 1 week. Initial interlaboratory frequency intercomparisons between cesium fountains, before regular calibration schedules can be set up, may require extrapolation times longer than 1 week for minimum uncertainty.

Envisaging multiple frequency calibration runs per week, of either hydrogen maser type with a cesium fountain having a standard uncertainty of $10^{-14}\tau^{-1/2}$ optimistically 5 per week, at the same time each working day, what is the best weighting procedure for using these calibrations in an algorithm to determine the frequency over a given interval? For the week's pattern, postprocessing extrapolation of each day's results independently, using the frequency from the nearest calibration interval gives a 77% improvement in accuracy for the free-running maser, and an improvement of 29% for the auto-tuned maser.

We have solved for the optimum weights of the maser calibrations to give the lowest standard uncertainty in average frequency over one week [6]. The week is best spanned by weighting Monday and Friday runs more heavily, to account for the weekend gap in calibrations. For the type 1 maser, the optimum weights follow the spanning times rather closely, and the optimum weights offer only a 1.1% improvement in average frequency. For a type 2 maser, there is a 4.7% improvement.

If adjacent weeks' calibration runs are also available, and the average frequency over a particular week is required, the optimum metafitting includes a small admixture from the preceding and the following weeks. For a type 1 maser, most of the weight comes from the preceeding Friday and the following Monday. For an autotuned (type 2) hydrogen maser noise model, the optimum weights have a slower variation through the weeks, and the three-week optimum has several % of the weight on points that are a full week from the calibration runs of the central week. There is a 19% improvement to the type 1 maser, and an 18% improvement for the type 2 maser. The improvements are summarized in Table I, given with standard uncertainties and cumulative advantages as each improvement is applied. For either maser model, the optimization of weights to apply to each run over multiple weeks gives about a 20% improvement in accuracy from the equal-weight case. It is not a large improvement, but it is almost free - although it does give additional cross-correlation between each week's frequency processed in this way. Cascaded with the other advantages discussed earlier, it results in a factor of 2.2 improvement in the accuracy transferrable with a free-running (type 1) hydrogen maser; and an improvement of 64% for the auto-tuned (type 2) maser. For the free-running maser model, the metafitted optimum standard uncertainty is 6.8 times smaller than

	method	Type 1 $u_y(7d)$	Adv.	Cum. adv.	Type 2 $u_y(7d)$	Adv.	Cum. adv.
I	f reset to unweighted fit	1.79×10^{-15}		1.00	1.15×10^{-15}		1.00
II	f reset to end points	1.77×10^{-15}	1.2%	1.01	1.15×10^{-15}	0%	1.00
III	phase lock + II	1.73×10^{-15}	2.3%	1.04	1.12×10^{-15}	2.4%	1.02
IV	daily postprocessed	0.97×10^{-15}	77%	1.83	0.87×10^{-15}	29%	1.32
V	metafit 1 week	0.97×10^{-15}	1.1%	1.85	0.83×10^{-15}	4.7%	1.39
VI	metafit 3 weeks	0.81×10^{-15}	19%	2.21	0.70×10^{-15}	18%	1.64
VII	metafit 5 weeks	0.81×10^{-15}	0%	2.21	0.70×10^{-15}	.1%	1.64
	$\sigma_y(7d)$	5.48×10^{-15}			1.72×10^{-15}		

Table 1: Reduction of standard uncertainty in average frequency at 7 days, for a free-running (type 1) maser, and an autotuned maser (type 2), when controlled by different methods from five 1-hour calibrations per week. The % advantage for each method is the accuracy improvement over the previous method. The last column gives each method's cumulative advantage over method I, a synthesizer reset to the least-squares calibration fit. The Allan deviation $\sigma_y(\tau = 7d)$ is also given.

the Allan deviation at 1 week, and for the type 2 maser it is 2.5 times smaller than the Allan deviation at 1 week.

Other interesting strategies are beyond the scope of this work. Longer runs on Monday and Friday and/or early-Monday and late-Friday calibration runs could be invoked to further improve the performance. Our methods allow for weight optimization for any set of calibration runs, and for calculating the resulting standard uncertainty in average frequency.

For some applications, statistical independence of each week, or each 10-day period, may be highly valued - for example, the clock reports to BIPM each 10 days that are used for determining TAI (and UTC) should be independent of each other. Weights for data from the weekly calibration cycle could be re-optimized for the seven different 10-day cycles that would exist. The metafitted optimum weights for the two maser models are shown in Figure 1. For the free-running maser model, the 70-day standard uncertainty in average frequency is 3.31×10^{-16} for the combination of the seven independent optimized 10-day periods, as compared to 3.07×10^{-16} for the combination of 10 independent 7-day periods. For the autotuning maser model, the 70 day standard uncertainty in average frequency is 2.72×10^{-16} for the combination of the seven independent optimized 10-day periods, as compared to 2.62×10^{-16} for the combination of 10 independent 7-day periods.

Conclusion

Our method for calculating the standard uncertainty for realistic noise models has allowed us to compare a wide variety of algorithms for treating one particular calibration schedule. We have metafitted the algorithm in several ways, and have identified ways to improve the accuracy of the maser frequency control by 2.2 and 1.64 times. We find that using the 10-day BIPM schedule, with independent processing of the calibrations for the 10-day periods, the expected asymptote for a single auto-tuned (type 2) maser could reach 1.2×10^{-16} at 1 year. For a free-running (type 1) maser, the standard uncertainty at 1 year would be 1.5×10^{-16}. Thus a flicker floor and accuracy of 10^{-16} for the cesium fountain is accessible for periods of a year with current masers carrying the time scale. Operating the masers at the stability level of the masers presents a challenge. Transferring 10^{-16} frequency accuracy to a second laboratory also presents a challenge. The reliability of a cesium fountain which might do this seems to be a major challenge, perhaps comparable to the challenge of making a cesium fountain with a flicker floor and accuracy of 10^{-16}. Perhaps the greatest value of this metafitting procedure is to show the very best performance which might be extracted from masers represented by these models. If greater accuracy is desired, then different approaches must be used.

References

[1] **Guide to the Expression of Uncertainty in Measurement**, International Organization for Standardization, Geneva, 1993. First edition. (Published in the name of BIPM, IFCC, ISO, IUPAC, IUPAP and OIML.)

[2] A. Papoulis, **Probability, Random Variables, and Stochastic Processes**, McGraw-Hill, New York, 1965. First edition pp. 400-402. (Omitted from subsequent editions.)

[3] R.J. Douglas and J.-S. Boulanger, **Local Oscillator Requirements for Timekeeping in the $10^{-14}\tau^{-1/2}$ era**, *Proceedings of the 1992 IEEE Frequency Control Symposium*, pp. 6-26 (1992).

[4] D. Morris, R.J. Douglas and J.-S. Boulanger, **The Role of the Hydrogen Maser Frequency transfer from Cesium Fountains**, *Japanese Journal of Applied Physics*,**33**, pp. 1659-1668 (1994).

[5] R. J. Douglas., J.-S. Boulanger and C. Jacques, **Accuracy Metrics for Judging Time Scale Algorithms**, *Proc. 25th Annual PTTI Applications and Planning Meeting*, pp. 249-266 (1993).

[6] J.-S. Boulanger and R.J. Douglas, **Frequency Control of Hydrogen Masers Using High Accuracy Calibrations**, *Proceedings of the 1994 IEEE Frequency Control Symposium*, pp. 695-708 (1994).

[7] D.W. Allan, **Time and Frequency (Time-Domain) Characterization, Estimation, and Prediction of Precision Clocks and Oscillators**, *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.* **UFFC-34**, pp. 647-654, (1987).

[8] J.-S. Boulanger, D. Morris, R. J. Douglas. and M.-C. Gagné, **Hydrogen Masers and Cesium Fountains at NRC**, in *Proc. 25th Annual PTTI Applications and Planning Meeting*, pp. 345-356 (1993).

[9] J.A. Barnes, A.R. Chi, L.S. Cutler, D.J. Healey, D.B. Leeson, T.E. McGunigal, J.A. Mullen Jr., W.L. Smith, R.L. Sydnor, R.C. Vessot and G.M.R. Winkler: **Characterization of Frequency Stability**, *IEEE Trans. Instrum. Meas.* **IM-20**, pp. 105-120 (1971).

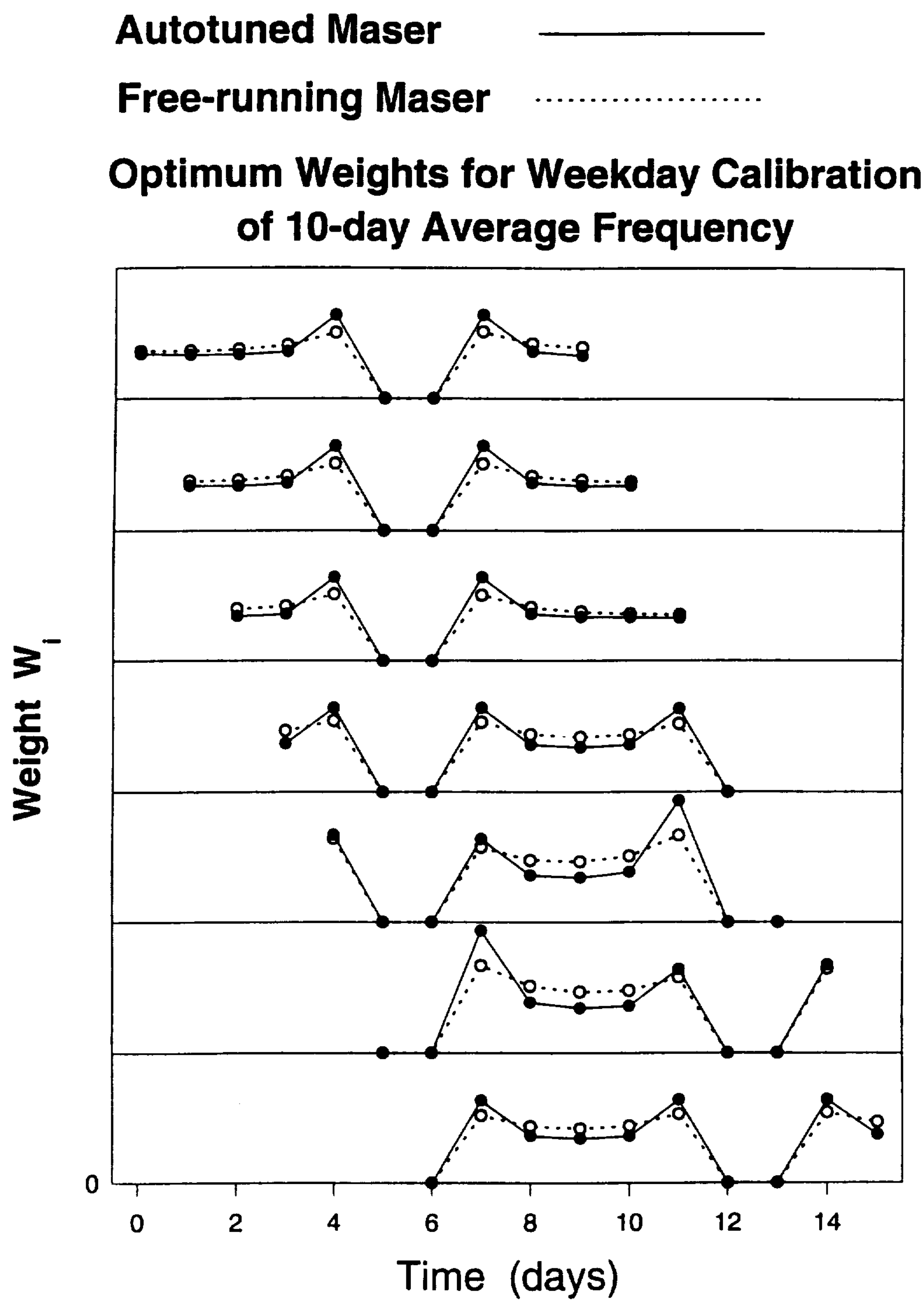

Figure 1. Optimum weights for combining weekday calibrations that give the minimum standard uncertainty in the average frequency over a 10-day interval. The calibrating reference standard is taken to be an ideal one, used for one hour, at the same time every working day. The optimum weights are shown for two flywheel oscillators: a free-running hydrogen maser model and an auto-tuned maser model. The optimum weights are shown for a the 10-day period starting on each day of the week.