Middle Temperature Error

Eur Ing David Boettcher CEng MIET

Introduction

Improvements in the timekeeping of marine chronometers ran into a problem at the end of the eighteenth century. The effect of variations in temperature on timekeeping had been almost eliminated by the use of compensation balances, but it was found that chronometers fitted with these could only be brought to time at one or two temperatures.

It was found that a chronometer fitted with a compensation balance which was brought to time at a certain temperature would lose at all temperatures above or below this. If it was made to gain at the this temperature it would then be correct at two temperatures, one above and one below. In the range between those two temperatures it would gain, the gain peaking at the middle temperature. At temperatures above or below this range it would lose. This phenomenon became known as "Middle Temperature Error" or MTE.

In this paper I don't attempt to discuss the history of marine chronometers. I will first briefly explain the two principal ways in which MTE was understood and explained, and then introduce a spreadsheet model that allows the user to explore the effects of the two different explanations.

Explanation 1: squares and square roots

The period "T" of a watch with an oscillator comprising a balance and balance spring is determined by the rotational inertia of the balance "I" and the couple exerted by the balance spring "S".

$$T = \pi \sqrt{\frac{I}{S}} \qquad (1)$$

A carbon steel balance spring gets weaker as temperature increases, and the balance also expands slightly. The weakening of the spring is bar far the great effect. If nothing were done the watch would run slower as the temperature increased.

A compensation balance as shown in Figure 1 has split bimetallic rims that compensate for the weakening of the balance spring by moving masses inwards to reduce the rotational inertia of the balance. The opposite effects happen if the temperature decreases. The angular position of the masses is adjusted to alter the amount of compensation, the distance that they move in or out in response to a change in

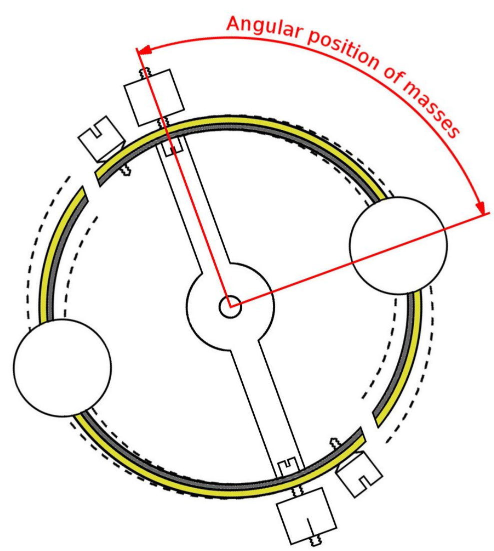

Figure 1: Compensation balance

temperature. The screws at the end of the cross bar are the “mean time screws” which are adjusted to bring the watch to time.

Although it had been observed since the late eighteenth century, MTE was first described in print, though not explained, by the chronometer makers Arnold & Dent in 1833[1]. An explanation for MTE was published in The Nautical Magazine of 1842[2] by Edward J. Dent of Arnold & Dent. A note in Shadwell[3] says that this explanation had been given to Dent by the Reverend George Fisher, although Fisher’s name does not appear in the article. Fisher had got into hot water with the Astronomer Royal by publishing a paper suggesting that the rate of ships chronometers was affected by magnetism in iron ships. The Astronomer Royal, Sir George Airy, refuted this in such strong terms that Fisher, although he continued to work on errors of chronometers, never published anything more. This is possibly why Fisher’s explanation of MTE was published under Dent’s name.

Dent gave the results of a test that he had conducted on a chronometer fitted with a glass disc balance, which indicated that the tension of the balance spring varied nearly linearly with temperature. He then illustrated the effect on the rate of a chronometer with Figure 2. The tension of the balance spring is represented by the straight line G G′ G″. The chronometer is regulated to mean time at B and B″.

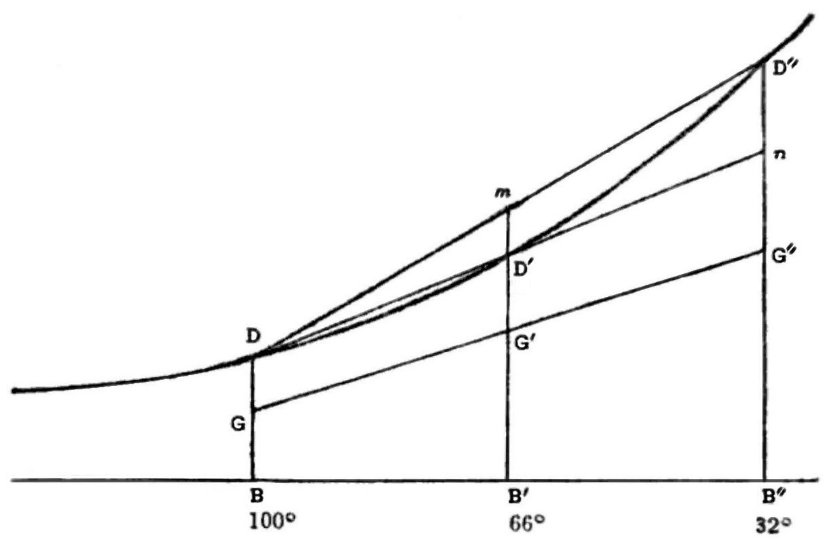

Figure 2: Dent’s illustration

The rotational inertia of a balance is given by $I = mk^2$, where m is the effective mass of the balance and k is the radial distance of that mass from the axis of rotation, usually taken for simplicity as the radial position of the compensation masses. This means that as the compensation masses move in response to changes in temperature, the effect on the inertia of the balance will be proportional to the square of k.

1 Arnold and Dent, The Nautical Magazine, London, Simpkin Marshall and Co. 1833, p. 224.

2 Dent, E. J., On the Errors of Chronometers, and a New Construction for the Compensation Balance, The Nautical Magazine, London, Simpkin Marshall and Co. 1842, p 760.

3 Shadwell, C. F. A, Notes on the Management of Chronometers (new ed), London, J. D. Potter, 1861, p18.

Dent represented the inertia of the balance as it changed with temperature as the curve D D′ D″ on the figure. The point D′ represents the inertia of the balance at the intermediate temperature B′, but the point "m" is what the inertia would need to be if the ratio of spring tension to inertia was to be the same as at B and B″ for the chronometer to go at mean time. Since D′ is less than it needs to be, the chronometer will gain, and will gain at all points where the inertia curve is below the straight line linking D and D″.

This explanation was widely accepted. It is related by Commander Rupert Gould[4] and illustrated with Figure 1, his figure 64. Like Dent, Gould considered the relationship of I to S given inside the square bracket of equation 1 above. The tension of the balance spring is represented by the straight line labelled the "S line", the varying inertia of the balance by the curves labelled "I curves".

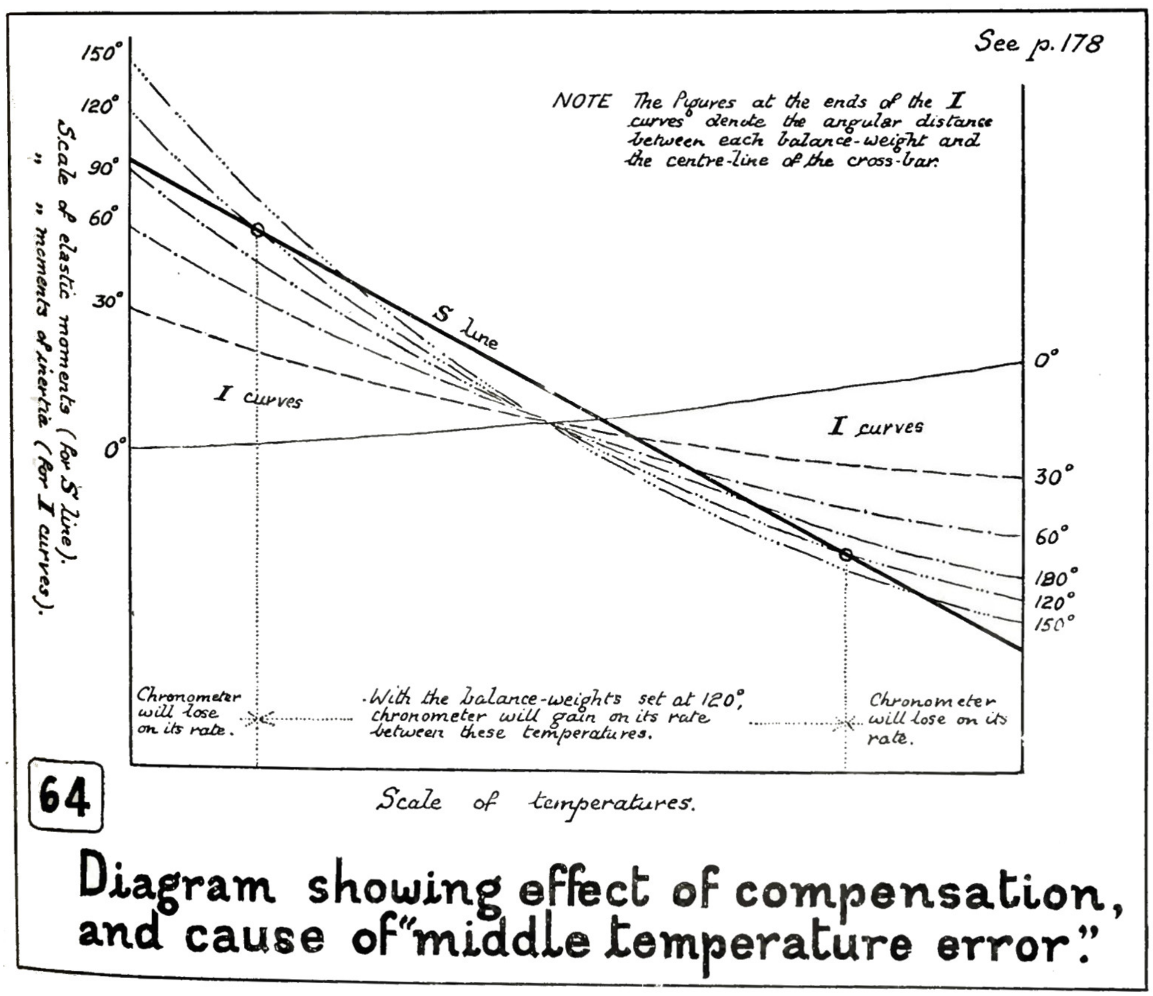

Figure 3: Gould's illustration

4 Gould, R. T. with foreword by Betts J., The Marine Chronometer: Its History and Development, Antique Collectors' Club Ltd, 2013.

The same explanation was repeated by Rawlings[5], with the difference that he considered the full equation for the period given in equation 1, including the square root, and not just the I and S terms inside the square root. Since rotational inertia is given by mk^2, when the enveloping square root sign is taken into account the inertia of the balance varies linearly with changes in the radial position of the compensation masses, but in inverse proportion to the square root of the couple of the spring. Figure 4 is Rawlings' plot the radial position “k” of the masses as a straight line against the spring couple, which he calls “Q”, as a square root curve.

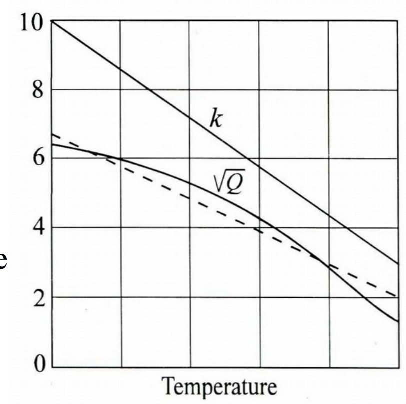

Figure 4: Rawlings’ diagram

Explanation 2: Material properties

In the early twentieth century, Charles-Edouard Guillaume[6], the discoverer of Invar, gave another explanation for MTE, which he referred to as “secondary errors”, and also sometimes as “Dent’s error”.

Guillaume’s explanation is illustrated by Figure 5. The horizontal axis labelled θ represents increasing temperature. The vertical axis represents changes in timekeeping, upwards for gaining and downward for losing.

The curves OL and OA represent the thermal expansion of the two parts of the bimetallic balance rims, L for the brass (laiton) and A for the steel (acier). These are curved because of non-linear effects in their thermal expansion. The dotted line OB represents the effect on the radius of gyration of the balance of the expansion of the two metals fused together. Guillaume explained that this is straight, because the quadratic coefficients of brass and steel are virtually the same and cancel each other out. This means that the compensation masses are moved in or out in a direct (linear) response to changes in temperature.

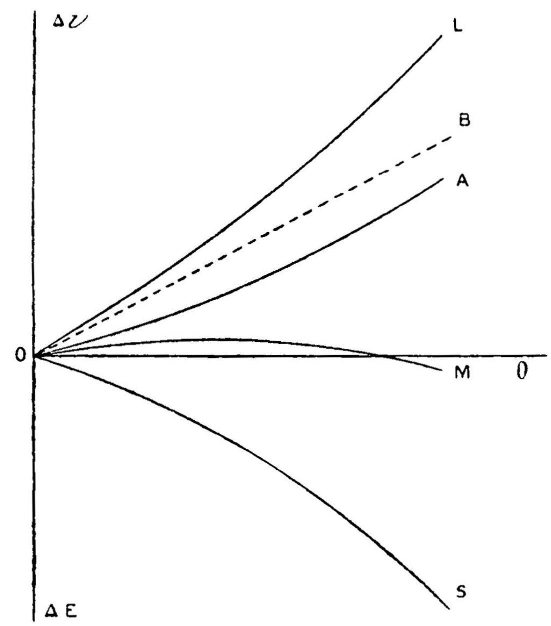

Figure 5: Guillaume’s explanation

The diagram is purely illustrative and is not to a consistent scale. The dotted line OB represents the difference between the expansion of brass and steel and should be well below the line OA. But if it was placed there it would not visually balance the line OS to produce OM.

5 Rawlings, A.L., The Science of Clocks and Watches 3rd ed, Timothy Treffry (ed). Upton, BHI, 1993.

6 Guillaume, C. E., Les Aciers Au Nickel, Annex to Horlogerie Théorique vol. 2, Grossman H and J., 1912.

When I first saw this diagram I thought that the straight dotted line OB could not possibly arise from the curves OA and OL so I plotted the three lines using data given by Guillaume. The result is shown in Figure 6. The difference between the curves of expansion of brass and steel is indeed the straight line towards the bottom of the plot.

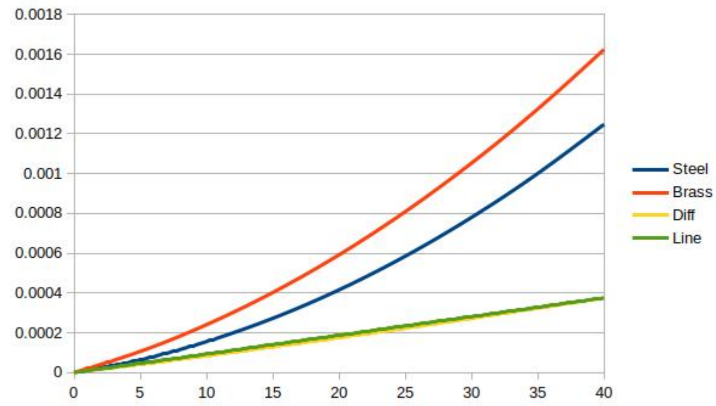

Figure 6: Differential expansion of brass & steel

The line OS on Figure 5 represents the decrease in the force of the spring. This line is also curved because of non-linear effects. Guillaume must have deliberately represented the dotted line OB out of scale so that he could draw the OS line to a similar scale in order that the curvature in the OS line would be visible.

The line OM on Figure 5 represents the sum of the effect of the balance inertia OB and the spring couple OS and shows the secondary error or MTE.

Guillaume invented in 1899 a modified form of the split bimetallic compensation balance to counter MTE. He used a nickel-steel alloy with a negative coefficient of thermal expansion that made the line OA curve down. This resulted in the line OB taking a curved path that mirrored the shape of OS. This balance was called the “Anibal” or “integral”. MTE was almost completely eliminated in chronometers equipped with these balances.

Which explanation is “correct”

Guillaume's explanation of MTE appears to have been virtually unknown or ignored in England, although it was correctly reported by Haswell[7] in 1928. Perhaps this was because Guillaume did not mention the square/square root effect, which had provided an intellectually satisfactory explanation for many years. Gould was clearly aware of Guillaume's work but does not mention the explanation of MTE being due the elasticity of the balance spring varying in a non-linear way with temperature. Instead Gould explains that Guillaume's balance counters the square/square root effect, and this explanation is echoed by Rawlings.

The square/square root effect explanation was widely accepted as the sole cause of MTE, at least in the English speaking countries, until the late Peter Baxandall noted that it couldn't fully account for the magnitude of observed MTE. Philip Woodward made a note to this effect in the BHI reprint of Rawlings, and went into more detail in an article in the Horological Journal[8] to which Jonathan Betts kindly drew to my attention.

7 Haswell, J. E. Horology (1928, supp. 1951), Wakefield, EP Publishing Ltd., 1976. p 159.

8 Woodward, P., Middle Temperature Error, The Horological Journal, Upton, BHI, April 2011.

Peter Baxandall must have realised that the curved lines drawn by Gould and Rawlings were exaggerated. The thermal coefficients of expansion and Young's modulus are very small, parts per million per degree C. And when coefficients are very small, their square roots do not behave as one might instinctively believe.

This peculiar behaviour of numbers close to unity was introduced to me quite casually in 1975 by my new boss on my first day in work. He was showing me a calculation for modelling hot CO_2 flow in a gas cooled nuclear reactor and we got to a point where the square root of 1.02 was required. This was at a time when pocket calculators were rare, but I had recently built one from a kit that I was rather proud of. Knowing that square roots are tricky I was reaching for my calculator when my boss simply wrote the answer as 1.01. This surprised me, I can still remember it clearly now. If it has a similar effect on you, just take a calculator and find the square root of 1.02. In the days before calculators were readily available, anything that simplified things was welcome, so it was useful to know that;

If $x \ll 1$, then $\sqrt{(1 + x)}$ is approximately $1 + x/2$.

This approximation gets more accurate as x gets smaller, because of course the square root of one is one itself. The important thing to notice is that 1 + x/2 is not a power expression but linear. The reason that this is important in the context of MTE is because we are dealing with small coefficients. For example, the temperature coefficient of Young's modulus quoted by Rawlings is 240 parts in a million per degree C. This means that the elastic modulus of a balance spring subject to a temperature range of +/- 20 degrees C will vary by a factor of 1.0048.

This is where the illustrations of Dent, Gould and Rawlings fall down. Although the charts of Dent and Gould are unscaled, Rawlings' chart of k and √Q has a scale and it is totally unrealistic. Instead of plotting the square root of a realistic range of values of the variations in Q, which for a temperature range of +/- 20 degrees centigrade would ranging from 1.0048 to 0.9952, Rawlings' chart shows the square roots of values from approximately 40 to 3. A nice curve for √Q is produced, which instinctively "looks right" but does not represent the true shape of the √Q curve in a real chronometer.

I discovered this when I tried to reproduce Commander Gould's elegant chart of I curves against a straight S line, using a spreadsheet and realistic values for the coefficients of expansion and elasticity. The I curves looked like straight lines, and the 120 degree I curve plotted exactly on top of the S line and obscured it. There is a curvature to the I curves of course, but it is extremely small and produce an MTE that is much less than observed values.

Realising that the square/square root explanation does not explain all of the observed MTE, although it undoubtedly does exist, Guillaume's explanation that the elasticity of the spring does not vary linearly with temperature becomes significant. In his HJ article Philip Woodward found that a value of 220×10^{-9} for the coefficient of curvature, called the "quadratic coefficient", in the elasticity of the balance spring was sufficient to explain the "missing" MTE. At first glance it seems that such a small number cannot have such a large effect. The spreadsheet that accompanies this article is intended to help the reader to investigate this effect.

Use of the spreadsheet

The spreadsheet is a very simple and theoretical model of a balance and balance spring using the methods explained in the appendix. It is not intended to model any real chronometer but simply to allow the effects of temperature on the balance and spring to be investigated.

The spreadsheet has three main areas. At the upper left is the “Input data” area. The cells outlined in green are the ones intended to be edited by the user. Below the input area the “Calculations” section shows the results of the calculations numerically. The two calculation sections show the changes in balance inertia I, spring torque S, and daily rate, over two ranges; plus and minus twenty degrees centigrade, and plus and minus one degree. To the right the results are displayed as two charts. The upper is a plot of I and S against temperature, imitating Commander Gould’s diagram. The lower chart is a plot of daily rate against temperature.

First steps

As supplied the spreadsheet is showing the effect of temperature on an uncompensated watch with a steel balance. The error for a temperature change of one degree C is just under 10 seconds per day.

To investigate the effect of a brass balance, the expansion coefficient for the balance should be changed to that of brass, i.e. 19.0 x 10^{-6}/ºC. The calculation in section 2.2 should then show the 10½ seconds per degree centigrade observed by Berthoud and analysed by Rawlings. Although the thermal expansion coefficients of brass and steel are considerably different, the effect on timekeeping is quite small. It is the variation in the elasticity of the balance spring that has the major effect.

Marine chronometer

To investigate temperature effects in a marine chronometer with a bimetallic compensation balance such as shown in Figure 1, first the expansion coefficient of the balance should be reset to steel, i.e. 11.0 x 10^{-6}/ºC. This determines the thermal expansion of the balance cross bar.

The "Compensation coefficient" is a measure of the rate at which the compensation masses are moved in or out by the bimetallic rim. This is not a material property but emerges from the design of the bimetallic balance itself, the diameter of the balance, thickness and proportions of the bimetallic rims, and the compensation masses. This is pre-set to 114.5 x 10^{-6}, a value determined using a method similar to that outlined by Philip Woodward to give optimum compensation for this model, much as the manufacturer of a real chronometer would have arrived at, no doubt by much sweat and tears.

The “Compensation mass angle” is the angular distance of the compensation masses from the centre line of the cross bar of the balance in degrees, as shown in Figure 1.

When the compensation mass angle is zero, the compensation masses are at the roots of the bimetallic arms, next to the cross bar, where they provide no effect for changes in temperature.

Increasing the compensation mass angle angle has the same effect as sliding the compensation masses around the rims, away from the cross bar and towards the free ends. The greater the angle, the further the masses are along the bimetallic rims, which increases the distance they move in or out for a certain change in temperature, and therefore the amount of compensation.

Commander Gould plots “I curves” for compensation mass angles of 30, 60, 90, 120 and 150 degrees. If you put these values, one after another starting with 30, into the green box to the right of “Compensation mass angle (degrees):”, the red "I curve" of the upper chart will incrementally swing round towards the yellow "S line", similar to Gould's illustration.

The I curve is calculated by the spreadsheet using the square of the radial position of the compensation masses, so it is truly a curve. But because of the small coefficient, it appears to be pretty well straight. When optimum compensation is achieved with the masses at 120 degrees, the red I curve disappears behind the yellow S line. This shows just how straight the I curve is, it can completely hide behind the S line!

As the compensation mass angle is increased the effect of the increasing compensation can be seen in the scale of the y axis of the plot on lower rate chart. At first it drops from hundreds to tens of seconds per day, and then when the compensation reaches the optimum it switches from a straight line to a curve, and the scale change to fractions of seconds per day. The curve shows the effect of the square of the radius of gyration in the balance inertia which is not visible in the I versus S plot, although it is really there.

As supplied the value for the “Mean time screws” is set to zero. The mean time screws can be “adjusted” by entering a value in seconds to bring the rate to zero, say at 5ºC and 35ºC – try starting at 0.15 and then tweaking the value slightly. This reveals the MTE effect, the rate gaining between 5ºC and 35ºC and losing outside that range, but the magnitude of the MTE is less than seen in practice.

As Guillaume noted, the change in the elastic modulus of steel with temperature means that the S line is not actually linear but has a curve. Curvature can be added to the S line by altering the quadratic coefficient in the upper green box.

I find that a value for the quadratic coefficient of 240.0×10^{-9} requires an adjustment to the mean time screw value to bring the rate to zero at 5ºC and 35ºC of about 2½ seconds per day, about what is found in practice. With this value the yellow S line becomes sufficiently curved that the red I curve becomes partially visible at either end.

This reveals graphically the answer to the question of which explanation is correct: they both are. The pure geometry square/square root explanation is intellectually satisfactory and correct, but the magnitude of the MTE it produces is small. The non-linearity in the spring curve is small, but it provides the balance of the MTE that is observed. In a compensation balance, both effects are at work.

Data sources

The coefficients of thermal expansion of brass and steel (19.0E-06 and 11.0E-06 respectively) and the linear coefficient of the elastic modulus, 240.0E-06, are from Rawlings. They are similar to other published data for steel, although figures vary depending on the composition and treatment of the steel. These values must therefore be taken as representative rather than exact.

I found that a quadratic coefficient for the Young's modulus of the balance spring of 240.0E-09 gives a middle temperature error of about 2.5 seconds per day. This is similar to the 220.0E-09 used by Philip Woodward in his HJ article, the small difference is because my model incorporates the effect of temperature on the balance spoke and the thickness of the spring. Published data on quadratic coefficients, especially in the room temperature range, is essentially non-existent, but these values are consistent with what data I have found. I asked the British National Physical Laboratory to help me find data for the quadratic coefficient of carbon spring steel. Their resident expert[9] told me that although data for the linear coefficients is readily available, data on quadratic coefficients is rare because it is very dependent on the exact material composition and treatment.

Dent's data

Dent conducted experiments on a marine chronometer fitted with a glass disc balance. These experiments have been said to show that the elastic modulus varied linearly with temperature but Dent was rather more circumspect, saying that "... the force of tension varies very nearly as the temperature within ordinary limits …" The results from these experiments, although only three data points, does exhibit a slight curve. I found that using a linear coefficient of 223.0E-06 and a quadratic coefficient of 690.0E-09 for the elastic modulus of the balance spring gives a curve that is good fit to Dent's data, which is of the same order of magnitude as Woodward's figure.

Real life

Both Dent and Gould considered whether the path taken by the compensation masses has an effect. It cannot be exactly radial as the spreadsheet model assumes. However, the masses do not move very far.

Britten[10] gives data for several marine chronometers. The balance diameters are around 1.2 inches or 30mm. Taking the compensation coefficient used in the spreadsheet model, 114.5×10^{-6}, the compensation masses of such balances would move approximately 0.0017mm in or out per degree centigrade of temperature change.

For a temperature change of 20 degrees centigrade this would result in a total movement in or out of about 0.034mm. For comparison, the thickness of a human hair is around 0.06 to 0.08 mm. With such a small movement I wouldn't think that the exact path has much effect. This also gives an insight into how exacting the construction and adjustment of such chronometers must have been.

9 Morrell, R. Private communication, National Physical Laboratory Materials Division, Teddington, 2014

10 Good, R., Britten's Clock & Watchmaker's Handbook 16 ed., London, Eyre Metheun, 1978

Finally

Who first observed MTE? Guillaume says it was Berthoud. Commander Gould says that it can't have been Berthoud, because his chronometer didn't have a split bimetallic compensation balance and therefore could not experience the square/square root effect. But there can be little doubt that Berthoud had observed the MTE due to the non linearity in the variation of elasticity of steel with temperature as described by Guillaume, and therefore Berthoud *was* the first observer.

Appendix: Derivation of spreadsheet equations

The basic equation for the period of a balance controlled timekeeper is

$$T = \pi \sqrt{\frac{I}{S}}$$

Expanding this to show the individual components gives

$$T = \pi \sqrt{\frac{m k^2 12 l}{t^3 E b}}$$

Where m is the mass of the balance, k is its radius of gyration, l, b and t are the length breadth and thickness of the balance spring, and E is the modulus of elasticity, Young's modulus, of the balance spring material. The 12 is a constant arising from the beam theory of bending.

Some of these terms are not affected by changes in temperature, i.e. the numerical constants, the mass of the balance, and the ratio of the length to breadth of the balance spring (both will change but in the same proportion) so we can separate these from the terms that are affected by changes in temperature as follows

$$T = \pi \sqrt{\frac{m 12 l}{b}} \sqrt{\frac{k^2}{t^3 E}} \text{ or } T = const \sqrt{\frac{k^2}{t^3 E}}$$

Effects of temperature

There are two effects of temperature on the radius of gyration. Thermal expansion or contraction will change the length of the balance spoke or cross bar and thus the radius of gyration. In a temperature compensation balance, the radius of gyration will change as the bimetallic rims move the compensation masses in or out. These effects can be represented for a temperature rise of h degrees as

$$k_h = k + k \alpha_b h - k c h \quad \text{or} \quad k_h = k\left(1 + \alpha_b h - c h\right)$$

Where "α_b" is the coefficient of thermal expansion of the balance spoke material and "c" is the coefficient of compensation, the rate at which the compensation masses move in or out in response to changes in temperature.

The spring thickness will vary as

$$t_h = t\left(1+\alpha_s h\right)$$

Where "α_s" is the coefficient of thermal expansion of the balance spring material.

The modulus of elasticity decreases with temperature. For many years this was thought to be linear, but it is known that it actually follows a curve. This can be represented by

$$E_h = E\left(1 - b h - q h^2\right)$$

where "b" is the thermal coefficient of the modulus of elasticity as discussed by Rawlings and others.

The curve in the thermal response of the coefficient of elasticity can be represented by the second order term in equation 6. The quadratic h^2 term gives a curvature, the quadratic coefficient "q" determines the strength of the curve.

The modulus of elasticity decreases with increasing temperature, hence the negative signs, and the rate of decrease increases as the temperature increases.

The change in period due to a temperature change h can therefore be written as

$$T_h = const\sqrt{\frac{\left(k\left(1+\alpha h - c h\right)\right)^2}{\left(t\left(1+\alpha h\right)\right)^3\left(E\left(1-bh-qh^2\right)\right)}}$$

The initial value of T gives 86,400 seconds in 24 hours. The ratio of T_h to T determines how many seconds the period T_h will make in 24 hours. I have not shown it in full, but the ratio of T_h to T can be shown to be

$$T_h/T = \sqrt{\frac{\left(1+\alpha h - c h\right)^2}{\left(1+\alpha h\right)^3\left(1-bh-qh^2\right)}}$$

The change in period of a watch that is keeping time at a certain temperature, when subjected to a change in temperature h, is determined purely by the thermal coefficients of the components that change their size or strength as a result of that temperature change. This follows the method of Philip Woodward but I have kept the effects of the thickening of the balance spring and weakening of its elasticity separate.

Once the change in period is known, what this means in terms of seconds per day lost or gained can be calculated, even without knowing the actual period.

The initial period of the watch when it has been brought to time must result in its counting 86,400 seconds per day. If, for example, the period is shorter and the ratio works out to be 0.9995, then the watch will count 86,400 x 1/0.9995 = 86,443.2 seconds, a gain of 43.2 seconds in 24 hours.