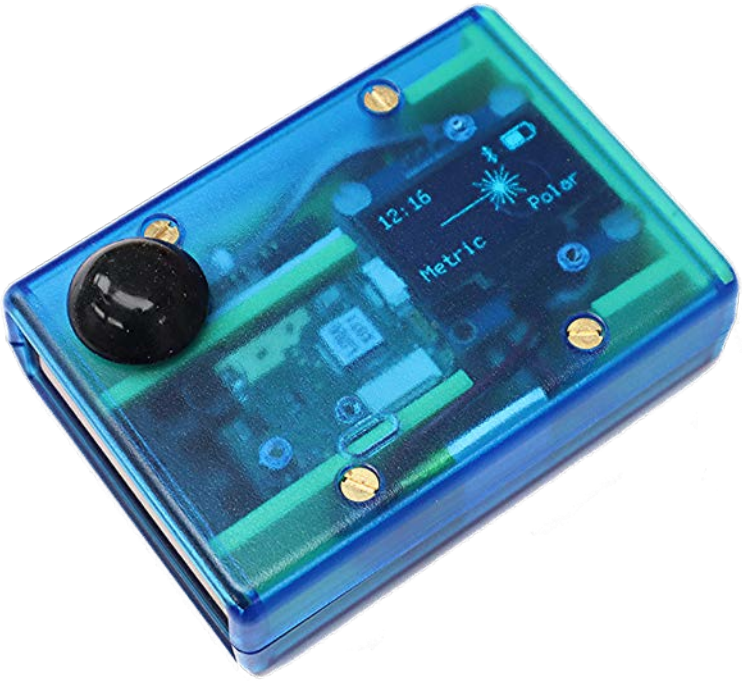
Non-linear Calibration of a Digital Compass and Accelerometer

***Phil Underwood** describes an approach to calibrating devices such as the Shetland Attack Pony (SAP), that addresses the non-linear response of sensors. This results in accuracy superior to more traditional methods.*

The Shetland Attack Pony (SAP) is a cave surveying tool which has undergone several iterations in its lifetime[1][2]. The original version was the first electronic compass/ clinometer developed for caving. In its current form, it consists of a 32bit CPU (PIC32MM0256GPM028 by Microchip[3]), a lithium-ion battery, an OLED display, a laser rangefinder, an accelerometer (LSM6DS3 by ST[4]) and a magnetometer (BM1422AGMV by Rohm[5]).

This is an open-source project – all the code and hardware designs can be found on GitHub[6]. This paper will present a method for increasing the accuracy of calibration by taking into account the non-linear response of the sensors used and will demonstrate the superiority of this method over more traditional calibration techniques.



The current iteration of the Shetland Attack Pony (SAP)

Principle of Operation

Laser rangefinder

The laser rangefinder uses a simple time-of-flight algorithm – it sends out pulses of laser light and records the time taken for those pulses to be received back. These are typically accurate to ±1 mm.

Magnetometer

The BM1422AGMV uses three orthogonal magneto-impedance sensors. These consist of a thin film of soft ferromagnetic materials with the special property that their impedance to a high frequency alternating current is highly sensitive to any external magnetic field. This component therefore generates a high frequency alternating current and measures the impedance to determine the magnetic field. This specific sensor is accurate to ±0.042 µT (Earth's horizontal magnetic field component is typically up to 30 µT, depending on location).

Accelerometer

The LSM6DS3 uses a micro-electro-mechanical system (MEMS) to measure acceleration. This consists of a small mass of silicon on a cantilever between two plates. Under acceleration, the mass moves towards one of the plates, altering the capacitance between the mass and the plate. The LSM6DS3 consists of three of these assemblies in an orthogonal pattern. It is accurate to ±90 µg (where 1g is 9.81 ms^{-2}).

Determining Orientation

We shall imagine the SAP initially to be lying flat on a plane with the display upwards. The y-coordinate extends in the positive direction parallel to the laser. The z-coordinate extends in a positive direction up through the display. The x-coordinate extends in a positive direction to the left when looking down at the display with the laser pointing away from you.

We can use the magnetometer and accelerometer to determine a vector for gravity $\boldsymbol{g}$ and magnetic field $\boldsymbol{m}$. From these we can calculate vectors for East and North using the cross-product.

$$\hat{\boldsymbol{e}} = \hat{\boldsymbol{g}} \times \hat{\boldsymbol{m}}$$

$$\boldsymbol{n} = \hat{\boldsymbol{e}} \times \hat{\boldsymbol{g}}$$

$$O = (\hat{\boldsymbol{e}}\ \hat{\boldsymbol{n}}\ \hat{\boldsymbol{g}})$$

The orientation matrix O then forms an orthogonal basis and we can use this to translate the y-vector (i.e. the direction of the laser beam (0,1,0)) into real world coordinates $\boldsymbol{y}$:

$$\boldsymbol{y} = O\begin{pmatrix}0\\1\\0\end{pmatrix}$$

$$azimuth = \text{atan2}(y_1, y_0)$$

$$inclination = \text{atan2}(\sqrt{y_0^2 + y_1^2}, y_2)$$

Traditional Calibration

The above calculations of course assume that we have perfect sensors, each exactly aligned, and that there is no magnetic interference from any other components on the PCB. Unfortunately, none of these assumptions are true. We can divide the potential sources of error into errors intrinsic to the sensors and extrinsic errors.

Intrinsic Errors

Each sensor has an offset $\boldsymbol{\beta}_{sensor}$ and a scale factor A_{sensor}, and also a degree of non-linearity. Non-linearity is the degree to which the sensor reading departs from a straight line, we shall denote it here as an arbitrary, unknown function $\psi(x)$, where x is the sensor reading.

$$\boldsymbol{x}_{sensor} = A_{sensor}\boldsymbol{x}_{raw} + \boldsymbol{\beta}_{sensor} + \psi(\boldsymbol{x}_{raw})$$

Extrinsic Errors

The incident magnetic field on the sensor can be affected by both hard iron

$\boldsymbol{\beta}_{sensor}$ (which is a simple offset) and soft iron effects $A_{softiron}$ (which can be a shear and scale transformation) within the device. Each sensor can also be rotated with respect both to the ideal orthogonal position, and also with respect to the device, so we also have a 3×3 rotation matrix R for each set of sensors.

$$\boldsymbol{x} = R\, A_{softiron}(\boldsymbol{x}_{sensor} + \boldsymbol{\beta}_{hardiron})$$

For the time being, we will ignore the non-linearity term, and we can simplify all this to

$$\boldsymbol{x} = A(\boldsymbol{x}_{raw} + \boldsymbol{\beta})$$

where A is a 3×3 matrix and $\boldsymbol{\beta}$ is a 3-vector.

Basic Calibration Process

We can make a first estimation at A and $\boldsymbol{\beta}$ by noting that both the gravitational and magnetic fields in any single location should be constant (at least over the time taken to calibrate the device). We take a series of 24 readings, of which 8 are distributed in multiple directions and orientations. The other 16 are two sets of 8 readings all taken from the same origin to two destination points, with the device held in varying degrees of roll along the y axis. This gives us a cloud of points which are distributed on the surface of an ellipsoid. We can use the process described in Vitali[7] to calculate the closest values for A and $\boldsymbol{\beta}$. This method is rather involved but essentially calculates the matrix and offset required that would translate all of the stored data points onto the surface of a unit sphere located at the origin, and minimises the average distance of any translated point from that surface. The error of this process can be estimated as below where N is the number of readings and $h_{magnitude}$ is the error.

$$h_{magnitude} = \sqrt{\sum_{i=0}^{N} \frac{(|\boldsymbol{x}_i| - 1|)^2}{N}}$$

Correcting for Rotation

Once this process is complete, we still have the possibility of errors due to misalignment between our sensors and the laser pointer. Essentially, we need to find the vector $\boldsymbol{v}$ in sensor coordinates that corresponds to the direction of the laser pointer. Consider the two sets of eight readings taken during calibration – for each set, the angle between the laser pointer and the magnetic/gravitational fields will be constant. They will therefore be located on a plane, which is normal to $\boldsymbol{v}$.

$$aX + bY + cZ = 1$$

$$D\begin{pmatrix} a \\ b \\ c \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\boldsymbol{v} = \begin{pmatrix} a \\ b \\ c \end{pmatrix}$$

We can use the equation for a plane to construct an equation where D is a 3×8 matrix consisting of the sensor readings from the particular set of readings. Solving this with a least-squares technique gives us a, b, and c, which can be used to construct $\boldsymbol{v}$. Once $\boldsymbol{v}$ is determined, it is trivial to construct a rotation matrix that can be applied to A to give us sensor readings in device coordinates.

We can also use these sets of readings to develop another measure of accuracy – essentially how similar the laser pointer vectors are in real-world coordinates:

$$h_{orientation} = \sum_{i=0}^{N} \frac{|\boldsymbol{y}_i - \bar{\boldsymbol{y}}|}{N}$$

$\boldsymbol{y_i}$ is the i^{th} reading of a set in real-world coordinates and $\bar{\boldsymbol{y}}$ is the average of all of those readings.

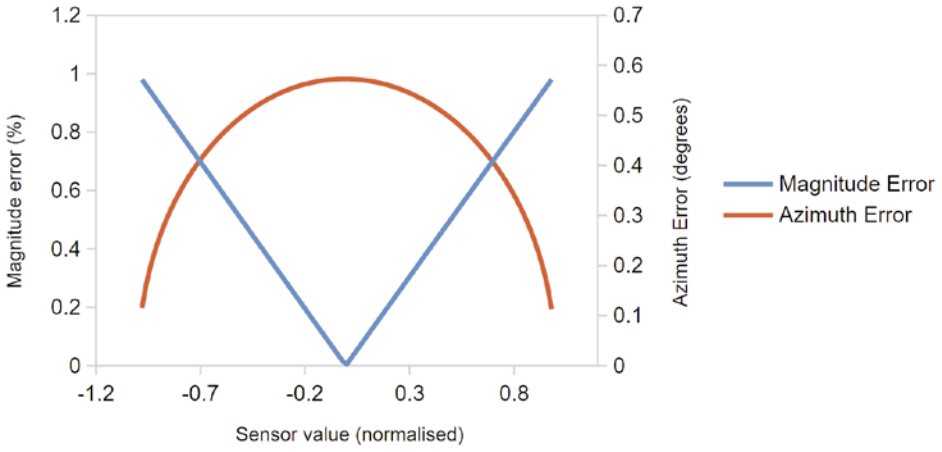


Figure 1 – Effect of Error on Magnitude and Azimuth

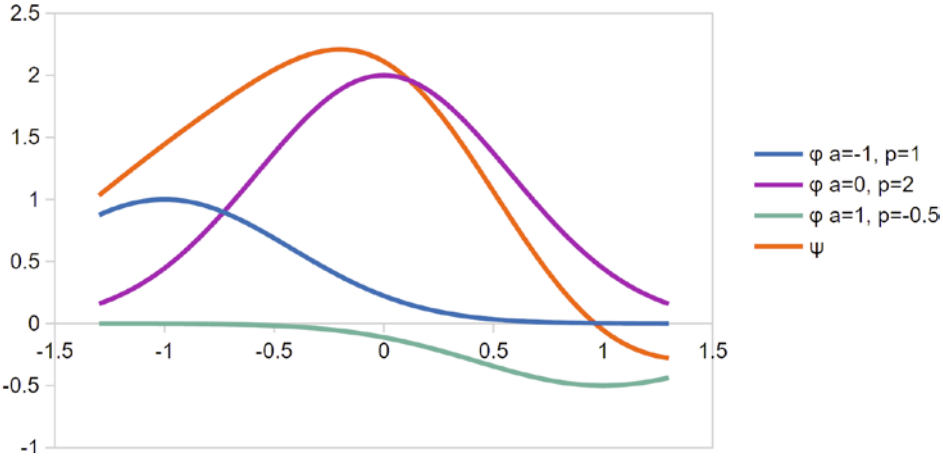


Figure 2 – Arbitrary Function from Radial Base Functions

Issues with Non-linearity

The previous calibration process worked on the principle that the output of the sensors varies linearly with the incident magnetic or gravitational fields. This is not always true, and the output of the BM1422AGMV in particular can have non-linearity of up to 2.8%. *Figure 1* shows the consequences of a 0.01 error (where 1 is the maximum reading when the sensor is aligned with the magnetic field). As we can see, the error in magnitude is greatest at larger sensor values, whereas the error in the azimuth (which is what we are interested in as cave surveyors) is greatest around the zero point. As the traditional calibration algorithm seeks to minimize errors in magnitude, then this can cause significant problems if the sensor response is not linear.

To progress, we need to find a way to determine the unknown function ψ(x). One common way to approximate an unknown function is to use polynomial fitting e.g.:

$$\psi(\boldsymbol{x}) = \sum_{i=0}^{P} p_i x^i$$

Where p_i is the i^{th} of P polynomial coefficients. Usually somewhere between two and ten coefficients are sufficient to approximate any arbitrary continuous function. Unfortunately, polynomial equations behave badly outside of the fitted range, which can cause problems if the device is moved to other locations with a stronger magnetic field. Instead, we can use a radial basis function as below.

$$\psi(\boldsymbol{x}) = \sum_{i=0}^{P} \phi(p_i, a_i, \boldsymbol{x})$$

$$\phi(p, a, \boldsymbol{x}) = p\, e^{-\left(\frac{(x'-a)P}{2}\right)^2}$$

$$a_i = \begin{cases} 0 & P = 1 \\ \dfrac{2i}{P-1} & P > 1 \end{cases}$$

This creates a set of P gaussian curves that can be added together to form an arbitrary function, but which rapidly tail off outside the interval (-1,1). $\boldsymbol{x'}$ is the raw sensor value, scaled and offset to fit in the range (-1,1); scale and offset factors can easily be calculated from A and $\boldsymbol{\beta}$. *Figure 2* shows an arbitrary function constructed from these radial basis functions. As you can see, these functions tend towards zero outside of their fitted range.

We can now use a minimisation algorithm to find the ideal values for p that minimises h (note that we still need to include $h_{magnitude}$ as otherwise the minimisation algorithm can find extreme values)

$$\boldsymbol{p} = \text{argmin}_{\mathbf{p}} h$$

$$h = h_{orientation} + h_{magnitude}$$

In principle, we could then repeat the ellipsoid fitting algorithm; however, in practice, this does not produce any substantial gains in accuracy.

Overfitting

One risk of approximating arbitrary functions is that of overfitting – where the generated function is determined more by noise in the data than by the underlying structure. As we increase the number of parameters for fitting the function, it starts to model the noise rather than the underlying function. We can assess for this using a 'leave-one-out' technique – essentially running the whole calibration process excluding the first data point and seeing how well that data point is placed using the calibrated constants, and then repeating this over all data points. This can be done using different numbers of parameters, and we can generally find a 'sweet spot' where we have enough parameters to model the non-linearity, but not enough to fall prone to overfitting. *Figure 3* shows several data points, which shows some data points from the cosine function with some added noise, with several polynomial approximations shown. The red line is underfitted with only 3 parameters, and misses the peak of the curve and also the tails on either side. The blue line is overfitted, and you can see it responding to noise in the data points, and also rapidly becoming unstable outside of the fitted range. The green line is 'just right' and models the general shape of the data without getting too distorted by noise.

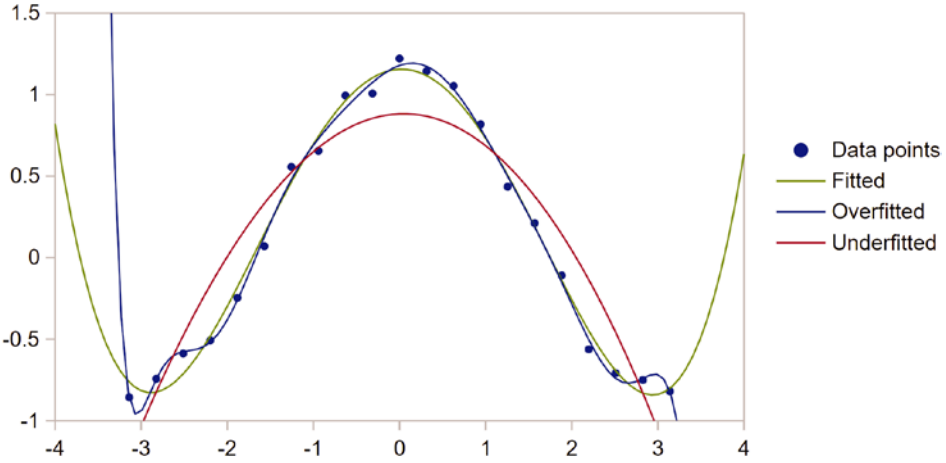


Figure 3 – Overfitting and Underfitting

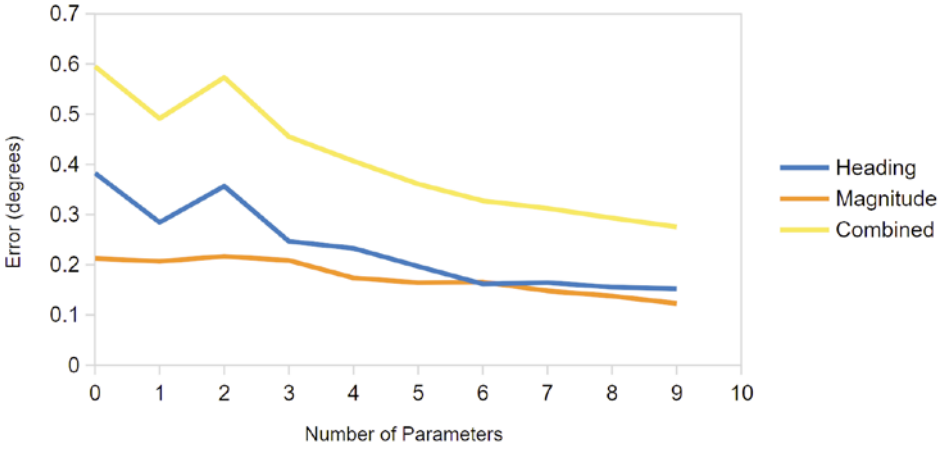


Figure 4 – Accuracy vs Number of Parameters per Sensor

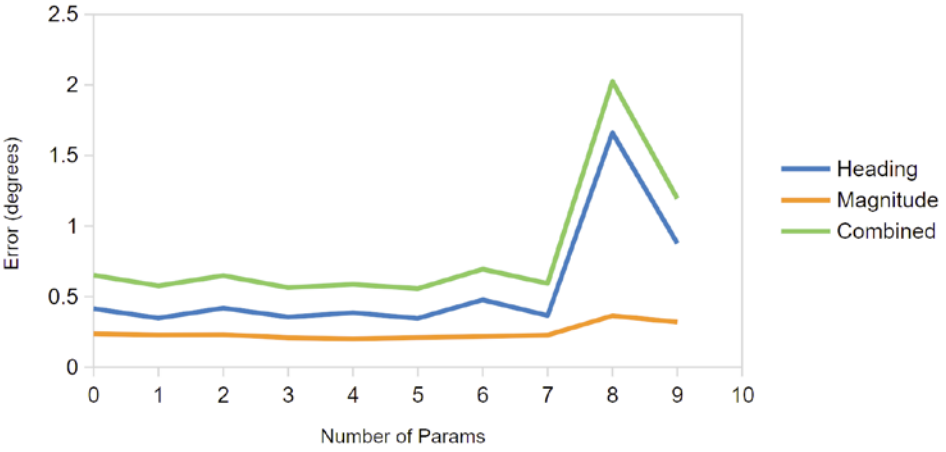


Figure 5 – Leave-One-Out Accuracy

Results

Nine SAPs were constructed and calibration data was obtained as described in the basic calibration process. The calibration algorithm described was then applied with between 0 and 9 parameters

for the non-linear correction (P). P=0 is equivalent to just using the 'traditional' calibration method. The results are shown in *Figure 4* and show that the majority of the benefit is achieved with three to five parameters, and then only gradually improves thereafter with more parameters.

The same procedure was also performed using a leave-one-out strategy to determine if there is any overfitting. The results are shown in *Figure 5* – they demonstrate good improvement in accuracy with up to five non-linear parameters. Using more than five parameters impaired performance due to overfitting, with severe problems when eight or more parameters are used.

Conclusions

This paper has demonstrated that non-linear effects in magnetometers can cause significant errors for electronic compasses that cannot be compensated with traditional calibration mechanisms. Where devices have a pointer, these errors can be minimised with the use of the technique demonstrated.

However, this technique can only correct for non-linearity in the x and z axes, as the y-axis sensor reading will not change substantially during rotation around the y-axis. It should be possible to use the flat bottom of a survey device as a way to rotate the device accurately around the z-axis, and this may provide a way to obtain a y-axis non-linear calibration – this is yet to be explored.

References

[1] Underwood, Phil (2007) *A Combined Electronic Compass and Clinometer*, CREGJ **66**, pp. 12–14, Mar. 2007.

[2] Underwood, Phil (2007) *Calibrating the Electronic Compass/Clinometer*, CREGJ **69**, pp. 10–13, Dec. 2007.

[3] Microchip Technology Inc., 'PIC32MM0256GPM028 - 32-bit PIC Microcontrollers'. **microchip.com/wwwproducts/en/PIC32MM0256GPM028** (accessed May 11, 2021).

[4] STMicroelectronics, 'LSM6DS3 - iNEMO 6DoF inertial measurement unit (IMU), for consumer electronics - STMicroelectronics'. **st.com/en/mems-and-sensors/lsm6ds3.html** (accessed May 11, 2021).[5] Rohm, 'BM1422AGMV - 3-Axis Digital Magnetometer IC | ROHM Semiconductor - ROHM Co., Ltd.' **rohm.com/products/sensors-mems/geomagnetic-sensor-ics/bm1422agmv-product** (accessed May 11, 2021).

[6] P. Underwood, SAP5 GitHub Repository. 2021. Accessed: May 11, 2021. [Online]. Available: **github.com/furbrain/SAP5**

[7] A. Vitali, 'Ellipsoid or sphere fitting for sensor calibration'. STMicroelectronics. Accessed: Apr. 17, 2021. [Online]. Available: **st.com/resource/en/design_tip/dm00286302-ellipsoid-or-sphere-fitting-for-sensor-calibration-stmicroelectronics.pdf**

Update: Energy Harvesting

*Further notes on energy harvesting and human-powered lighting by **David Gibson**.*

In the last Journal I wrote about energy harvesting (Gibson, 2021) and I made reference to Andy Lillington's human-powered torch (Lillington, 2020) where energy was obtained via the motion of a rocking magnet. I explained how his concept of a pendulum set the design apart from inferior designs that attempted to capture energy directly from the random shaking of a magnet. The salient point of Andy's design is that micro-movements of the magnet are first captured by a mechanical storage device – i.e. the pendulum – rather than going to waste as they would if they were too small to generate usable power by themselves.

Using a Cylindrical Coil

It has since occurred to me that it is not necessary to use a rocking magnet for this. It might be a design improvement to use a cylindrical induction coil and to have the magnet bounce up and down on a spring. Being entirely linear, such a device might be easier to construct. Of course, some attention needs to be given to the strength of the springs above and below the magnet, but the same principle would apply as it does to the pendulum, namely that the springs do not in themselves provide energy, they merely provide a means of capturing micro-movements and facilitating energy transfer.

Such a device might look like one of the shakeable torches now available – e.g. we reviewed the NightStar in CREGJ 66 (Bedford, 2007). However, the resemblance is only superficial because, for efficiency, the energy has to be transferred via an oscillating spring, which the NightStar does not have. The NightStar does have two magnets – one at either end – providing 'bounce' but this is reportedly to reduce noise, not to store energy. The salient point is that when stationary, the NightStar would not oscillate, unlike my proposed bouncing-spring device or Andy Lillington's pendulum.

Using a Switching Regulator

Another point that occurred to me was that the energy might be better harvested if a switching regulator was used to control the output voltage of the device.

Of the three 'elemental' form of switching regulator (boost, buck and inverting) it is the boost regulator that has the inductor on its input so the use of a boost regulator means that the harvesting inductor can double as the current regulating inductor. The output voltage is clamped by the LED so it is the input that we need to regulate, not the output; and the peak input current must be limited to the maximum pulsed current that the LED will tolerate.

Because the number of turns on the input inductor is related to the resistance and thus to the efficiency of the conversion, one might consider that a few turns of very thick wire might out-perform a design with a large number of very thin turns. Or perhaps, as in a cave radio antenna, the number of turns will prove to be irrelevant for a fixed mass of wire. There is some interesting analysis to be done here.

References

Bedford, Mike (2007, *NightStar - a Review*, CREGJ **66**, p11. March 2007.

Gibson, David (2021), *Energy Harvesting with Electrets*, CREGJ **113**, p16. March 2021.

Lillington, Andy (2020), *A Human-powered Torch*, CREGJ **109**, pp12-14. March 2020.