Development of a digital proton magnetometer: an efficient algorithm model to determine the Larmor precession frequency.

André Wiermann and Luiz C.C. Benyosef, OBSERVATÓRIO NACIONAL, Brazil.

Abstract

This paper presents results for an algorithm model proposed to estimate the central value for small measurement data sets. This algorithm was derived to achieve a computationally efficient method to estimate the Larmor precession frequency for a digital proton precession magnetometer that is being developed at the Observatorio Nacional (Wiermann & Benyosef, 2004). An additional comparison is shown between the proposed algorithm and a technique called *meadian* that combines mean and median procedures and it's based on statistical bootstrapping.

Introduction

Proton magnetometers measure the total geomagnetic field using atomic resonance. The sensor is a cylindrical container filled with a liquid rich in hydrogen atoms, usually kerosene, surrounded by a coil. The sensor is connected to an electronic unit with an electronic switch, an amplifier, and a frequency counter. When the switch is closed, a DC current is directed through the coil, producing a relatively strong magnetic field in the fluid-filled cylinder.

The hydrogen nuclei (protons), which behave like minute spinning dipole magnets, become aligned along the direction of the applied field. Power is then cut to the coil by opening the switch. Because the Earth's magnetic field generates a torque on the aligned, spinning hydrogen nuclei, they begin to precess around the direction of the Earth's total field. This precession produces a time-varying magnetic field, which induces a small alternating current in the coil. The frequency of the AC current is equal to the frequency of precession of the nuclei. Because the frequency of precession is proportional to the strength of the total field and because the constant of proportionality is well known, the total field strength can be determined quite accurately.

The classic estimation using arithmetic mean is based on the premise that the data set posses a normal (or near normal) and symmetrical distribution. A random noise that pulses with infrequent rates like electronic glitches or spikes on acquired signals, although normally distributed, tends to look like asymmetrical and non-normal when just a small number of samples is available. Under such situation the distribution function is unpredictive and usually the median becomes the estimator of choice.

Median has the advantage to be robust under the presence of outliers (Hoaglin et al., 1983) but have the problem to be not well behaved when data escapes from Gaussian (Wilcox, 2001). In general when data is not unimodal there is no meaning on central value estimation but this is not quite true for digital frequency counting.

Under certain circumstances it is possible to know at least in an approximate form the distribution of a given data collection. This is the case of digital counting in which main uncertainties are due to short term effects that makes readings to jump around a couple of adjacent counting. This behavior typically leads to a bimodal quasi-symmetrical distribution that sums to spurious counting (spikes) and other noise types induced from electromagnetic sources or generated inside the analog front-end electronics that typically precedes the digital counters (filters, Schmitt-triggers among other devices).

Method

When no initial indication about the distribution shape or function is given, specially with small data sets a procedure largely accepted is to determine that function as well as its variance from the data itself using the statistical technique called bootstrap (Efron, Tibshirani, 1993). The problem with bootstrap is its computational cost. From a set of data, several random subsets must be created using resample with repetitions taken from original data. For each subset a new variance must be calculated and a final combined variance can be extract along a new distribution plot.

Given a data set:

$$X = \{x_1, x_2, ..., x_N\} \quad (1)$$

Several subset samples ***B*** must be created randomly as:

$$B = \{X^{*1}, X^{*2}, ..., X^{*B}\} \quad (2)$$

With these new data sets we can calculate a large number of new mean and medians:

$$\{\overline{X}^{*1}, \overline{X}^{*2}, ..., \overline{X}^{*B}\} \quad (3)$$

and

$$\{med(X^{*1}), med(X^{*2}), ..., med(X^{*B})\} \quad (4)$$

Combination of mean and median in a robust way with the help of bootstrap can be made in a new proposed method called *meadian* (Josselin & Ladiray, 2002). Its shown that the meadian estimation stays most of time between the mean and median results, even under mild asymmetrical distributions, showing in general a good performance at some computational complexity cost.

Laplace defined a corrective term *C*, based on the covariance between mean and median to create a optimal linear combination of both:

$$C = \frac{V(\bar{x}) - Cov(\bar{x}, M)}{V(\bar{x}) + V(M) - 2Cov(\bar{x}, M)} \quad (5)$$

where $\bar{x}$ is the mean, *V* the variance and *M* is the median.

Covariance between mean and median can also be estimated by bootstrap:

$$\hat{C}ov(\bar{x}, M) = \frac{1}{B-1}\sum_{b=1}^{B}\left[\bar{X}^{*b} - m^{*}_{\bar{X}}\right]\cdot\left[med(X^{*b}) - m^{*}_{med(\bar{X})}\right] \quad (6)$$

The *meadian* method is based on a original work from Laplace, but using just mean and median variances as substitute for the more complex covariance:

$$\overline{M} = \frac{V(M)\bar{x} + V(\bar{x})M}{V(\bar{x}) + V(M)} \quad (7)$$

or,

$$\overline{M} = (1 - C)\bar{x} + CM \quad (8)$$

with,

$$C = \frac{V(\bar{x})}{V(\bar{x}) + V(M)} \quad (9)$$

The equations (7) to (9) represent a combination of mean and median weighted by the inverse of their variances (Josselin & Ladiray, 2002). This method while simpler than Laplace original is still heavy on processing due to bootstrap algorithm.

With the fundamental knowledge that a center value really exists (whatever kind of "center" it means) we can speculate about an algorithm to estimate it. When determining the frequency from a given signal using a digital counter, we can assume that the mean and median should produce approximate results depending on signal to noise ratio (SNR) and noise shape. Even though the meadian alone is more adequate for estimating digital counting, for simulation purposes we can apply a simple Gaussian noise with favorable SNR and use the results as a comparison starting point.

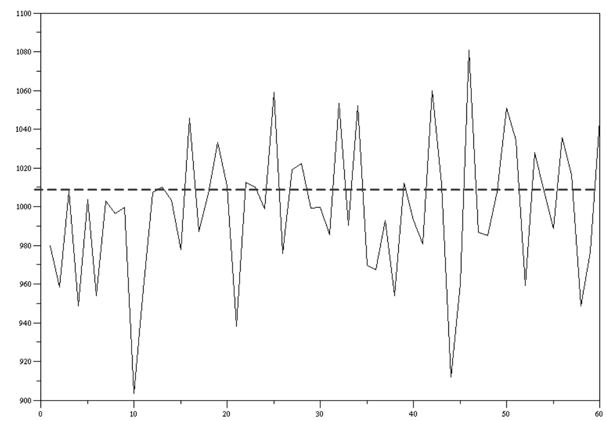
The usual procedure to calculate the median is to sort the values (classification) and pick up the middle point from resulting list. This procedure although simple and effective still leads to a biased estimation on very asymmetrical distributions. It is resistant to outliers but not efficient when exposed to some typical distribution laws.

A better way to improve its efficiency would be to take a weighted mean from a set of "median" data, with weights taken from each data deviation:

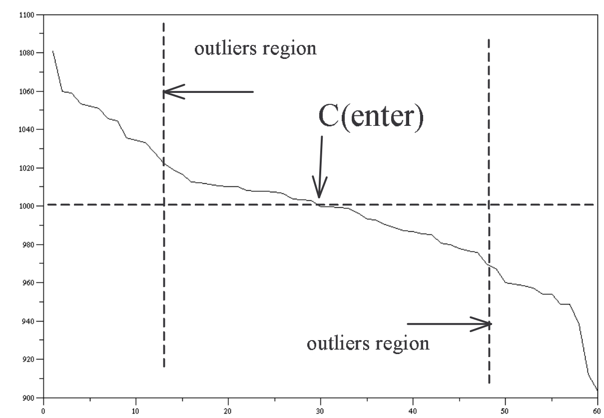
$$\overline{Med} = \frac{\sum_{n=(C-N/2)}^{C+N/2}\left[med_n \cdot \frac{1}{dev(med_n)}\right]}{N+1} \quad (10)$$

where *C* is the center index for the median and *N+1* is the window size used for data estimation.

The problem with this approach is that on determining the deviation for each data, a premise on the center *C* - the arithmetic average or equivalent - is required leading to a estimation biased towards that average. Another problem that persists is how to define the region where the weight will be applied (size of *N* and position of *C*).



(a)



(b)

Fig. 1 - Synthetic data (60 points) with strong noise around 1000Hz before **(a)** and after **(b)** sorting for weighted mean of median estimation. A set of numbers around the median or other center estimator can be taken to be averaged. As the distribution is almost symmetric the median would be a good estimator.

Proposed algorithm

The proposed algorithm is based on the fact that even for moderately asymmetrical distributions the sorted values should present a region (not necessarily at the center) that is smooth and flat or at least, not as steep as all other regions on the set (fig.2).

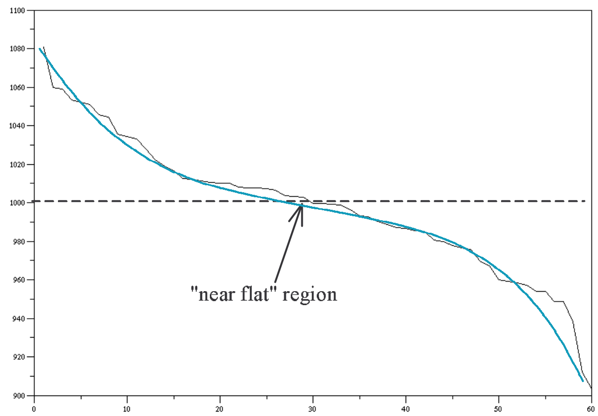


Fig.2 - Synthetic data (as in fig.1) showing the region where the samples go smooth and near flat (after sorting).

This place coincides with a peak on data histogram, representing a trend to a "central" value (or a escape from outliers). Such region (fig. 3) can be detected by a simple algorithm that looks for a minimum derivative (Fusett, 1999) within a given limit inside a window and defined by metrics for the specific application.

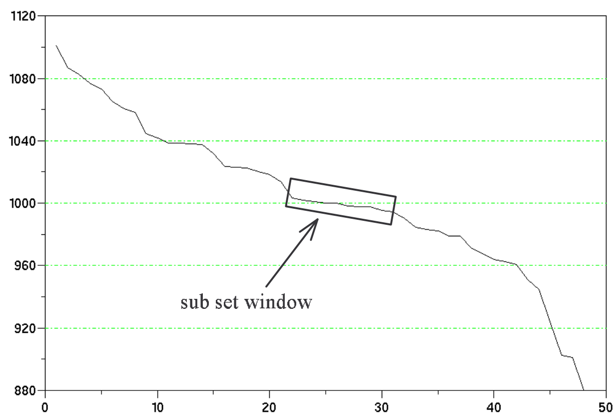


Fig.3 - Based on the data slope (smaller derivative inside a window) one region is selected to represent a sew data set for estimation.

After defining the window a simple mean or median can be used to pick up a final value without any significant difference in the resulting estimation. This method was implemented on a microprocessor using C language and embedded in a proton precession magnetometer with excellent preliminary results on bench tests (Wiermann & Benyosef, 2004).

Results

Before embedding the algorithm on the instrument, a large number of synthetic data using random values and spikes were used to evaluate and compare its estimation performance.

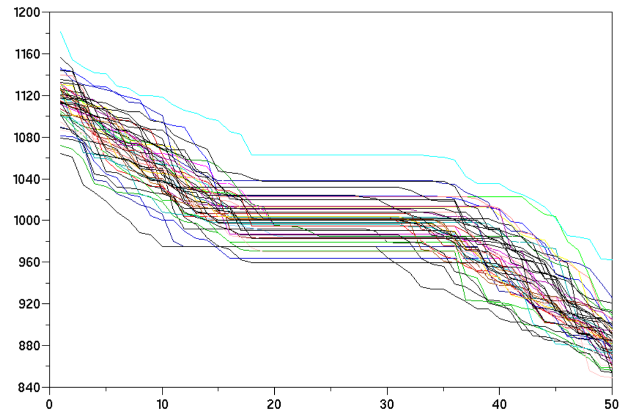


Fig.4 - 100 simulations of 1000Hz with Gaussian white noise and uniformly distributed Gaussian spikes (20dB total SNR). Data is shown after been sorted.

The figure 4 shows one simulation based on 100 runs whose synthetic value of 1000Hz was contaminated with Gaussian white noise and uniformly distributed Gaussian spikes. The SNR was 20dB, twice the level achieved by the proton magnetometer electronic filters.

As on each data measurement the results fluctuates for all estimators, a set of histograms were built to show after several runs how each technique (or algorithm) behaves.

Can be seen by the results on figure 5 that after several running as the total number of estimations grows the mean become centered on the ideal value. But if just a few measuring are available, the standard deviation and histogram shows that the proposed combination of mean with median gives more accurate estimations.

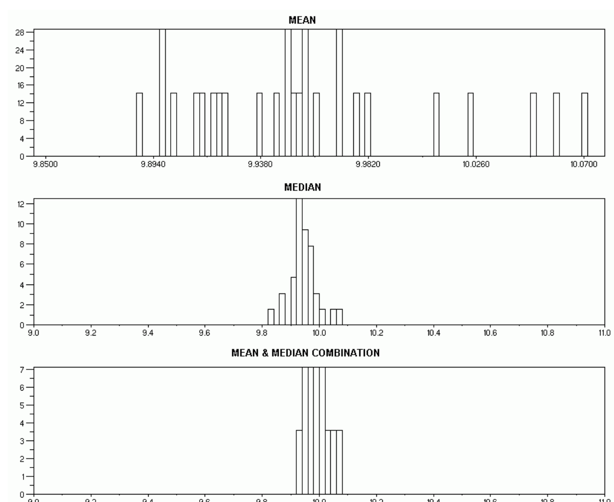


Fig.5 - Histogram for several tests comparing the mean. Can be seen from this graphic that as the data distribution was quite asymmetric and sparse the mean fluctuates along all tries while the median and the combination stay most of time near the ideal value.

Following on the table 1 is a summary with the several estimations and its standard deviations to show their relative performances:

Estimator	Error $(x-\hat{x})$	SD
Mean:	-2.699482	6.494555
Median:	-3.740123	6.494555
Combination:	-2.682701	6.070788

Table 1 - Final results showing the average and deviation for mean, median and the proposed combination.

Conclusions

The proposed algorithm is simpler and faster than some complex methods like "meadian" and Laplace linear mean/median interpolation. This method depends on the parameters (derivative window length and error limit) defined by own user to fit better on specific distributions. As an investigation tool, those parameters can be tested to find convergences that can indicates peaks on the data histogram. When applied to the strongly centered data contaminated by outliers it behaves similarly as the standard median.

Acknowledgments

To **FAPERJ** for the support to the LDSM/ON.

References

Efron B. and Tibshirani R. J., 1993 An introduction to the Bootstrap, Monographs on Statistics and Applied Probability 57, Chapman & Hall, New-York.

Hoaglin D., Mosteller F., and Tukey J.W., 1983 Understanding Robust and Exploratory Data Analysis, Wiley Series in probability and mathematical statistics, New-York.

Fausett L.V., 1999 Applied Numerical Analysis, Prentice Hall, New Jersey.

Huber P.J., 1983 Robust Statistics, Wiley Series in probability and mathematical statistics, New York.

Josselin D. et Ladiray D., 2002 "Combining L1 and L2 Norms for a more Robust Spatial Analysis: the Meadian Attitude", First Review - Presented in European Colloquium on Theoretical and Quantitative Geography, Saint-Valery en-Caux France, 2001.

Rousseeuw P.J. and Leroy A.M., 1987 Robust Regression and Outlier Detection, Wiley Series in probability and mathematical statistics, New York.

Wiermann, A. and Benyosef, L.C.C. 2004, PPM Development using distilled water as sensor liquid – IV Reunión Nacional de Ciencias de La Tierra – UNAM – Proceedings in CD ROM – Cidade do México

Wilcox, R.R., 2001 Fundamentals of Modern Statistical Methods, Substantially Improving Power and Accuracy. XIII, 258 p Springer-Verlag.