Dr. Manry Study-Coal Seam Thickness

The following study was done by Dr. Michael Manry at the University of Texas for a major coal mining company. The purpose of the study was to compute the probability of success when using the PetroSonde system for determining coal seam thickness and how to integrate simple correction factors with PetroSonde data. In fact, the probability computed to be in excess of 90%. The actual computations are available upon request.

Estimating Coal Seam Thickness Using Multiple Petro-Sonde Measurements

M.T. Manry E.E. Dept.
University of Texas at Arlington

Our goals here are to (1) describe simple methods for correcting Petro-Sonde-based estimates of coal seam thickness, and to (2) describe the calculation of the corresponding confidence intervals.

A. Confidence Intervals

Assume that *Core'* is an estimate of *Core,* which is the actual measured thickness of a coal seam. *Core'* may be calculated from any number of Petro-Sonde measurements, taken in a given area. The bias *b* of *Core* ' is defined as

$$b \equiv Core - E[Core']$$

where ED denotes the statistical expected value or average of the quantity in square brackets. The estimation error *e* is defined as

$$e \equiv Core' - Core$$

The random variable $(Core\text{-}Core'\text{-}b)/\sigma_e$ is zero-mean and has unit variance. We assume that this random variable is Gaussian. Let *F(x)* denote the probability distribution of a zero-mean, Gaussian random variable having unit variance.

The confidence interval for the estimate *Core'* is defined as

$$Conf(Core-b-k < Core' < Core-b+k) = \gamma$$
$$\equiv P(|e+b| < k) = \gamma/100$$

In other words, if we are γ % confident that *Core'* lies between *Core-b-k* and *Core-b+k,* this is equivalent to saying that the probability that $|e+b| < k$ is $\gamma/100$. Now,

$$P(|e+b| < k) = \gamma/100$$

can be rewritten as

$$P\left(\left|\frac{e+b}{\sigma_e}\right| < \frac{k}{\sigma_e}\right) = \frac{\gamma}{100}$$

$$1-2\cdot F\left(-\frac{k}{\sigma_e}\right) = \frac{\gamma}{100},$$

$$F\left(\frac{k}{\sigma_e}\right) = \frac{1+\frac{\gamma}{100}}{2}$$

$$F(c) = \frac{1+\frac{\gamma}{100}}{2}$$

Solving the last equation for c, we then get

$$k = \sigma_e \cdot c$$

In the table below, we list values of and the corresponding values of the constant c.

Table 1. Values of γ and c for Confidence Interval Calculation

γ	c
90	**1.645**
92.5	**1.78**
95	**1.96**
97.5	**2.24**
99	**2.576**
99.9	**3.291**

If *Core'* is calculated as the average of *M* measurements, each with a variance and σ_1 and a bias b_1 then the values of *k* and *b* are

$$k = \frac{\sigma_1}{\sqrt{M}} \cdot c,$$

$$b = b_1$$

Now, it remains for us to look at methods for calculating *Core'* from the Petro-Sonde core thickness estimates, *Etu.*

B. Calculating *Core'* Using A Correction Factor

Here, our goal is to develop a signal model for thickness estimates as

$$Core\ /\ (n)\ Core(n) + \mathrm{e1}\ (n)$$

where *Core(n}* denotes the thickness measurement, from a core, corresponding to the *nth* core estimate *Etu(n). Etu(n}* comes from an engineer operating the Petro-Sonde. The corrected core thickness estimate, *Core' (n),* is found as

$$Core'(n) \equiv B_1 \bullet Etu(n)$$

We have N measurements, where $N = 52$.

Our tasks are to

(1) Find the correction factor B_1 such that the mean-square-error (MSE)

$$E = \sum_{n=1}^{N} [\ Core(n) - B_1 \cdot Etu(n)\ ]^2$$

is minimized. This estimate of B_1 is called the minimum MSE estimate.
(2) Find the estimated mean m_1 of the noise $e_1(n)$ and the bias as

$$m_1 \equiv \frac{1}{N} \sum_{n=1}^{N} (\ B_1 \cdot Etu(n) - Core(n)\ ),$$

$$b_1 = -m_1$$

(3) Find the estimated noise variance per sample,

$$\sigma^2_1 = \frac{1}{N} \sum_{n=1}^{N} (e_1(n))^2 - (m_1)^2$$

Taking the partial derivative of E with respect to B_1 and equating it to zero, we get

$$B_1 = \frac{\sum_{n=1}^{N} Core(n) \cdot Etu(n)}{\sum_{n=1}^{N} (Etu(n))^2}$$

Using this equation on our 52 data points, we get

$$B_1 = .80462$$

thereby satisfying task (1). Performing tasks (2) and (3), we get

$$\sigma_1^2 = 1.055697,$$

$$\sigma_1 = 1.027471,$$

$$m_1 = -.14765$$

1. Estimating Seam Thickness from Multiple Measurements

Assume that we have acquired *M* seam thickness estimates, *Etu(n)* for $1 \leq n \leq M$, and corrected them with the factor B_1. Assume further that all the data was acquired in a small area, where the true seam thickness, called *Core,* is constant. We now develop a method for calculating *Core',* an estimate of *Core,* from the *M* values *Core' (n),*

In our signal model,

$$Core'(n) = Core + e_1(n)$$

the average value or mean of $e_1(n)$ is *m1,* The likelihood function for estimating *Core* is

$$p(Core' \mid Core) = \frac{1}{(2\pi)^{\frac{M}{2}} \sigma_1{}^M} \cdot e^{\frac{-1}{2\sigma_1^2} \sum_{n=1}^{M} (Core'(n) - Core - m_1)^2}$$

The corresponding log likelihood function is

$$LLF \equiv \ln(p(Core' \mid Core)) = C - \frac{1}{2\sigma_1^2} \sum_{n=1}^{M} (Core'(n) - Core - m_1)^2$$

Taking the partial derivative of LLF with respect to *Core* and setting it equal to zero, we get

$$Core' = -m_1 + \frac{1}{M}\sum_{n=1}^{M} Core'(n)$$

and the variance of the noise, *e,* in *Core*' is

$$\sigma_e^2 = \frac{\sigma_1^2}{M}$$

Note that m_1, the mean of the noise for one measurement, is subtracted off the average value of *Etu(n),* in the optimal estimate above. This step may be unacceptable in practice, since it results in a positive value for *Core'* even when *Etu(n)* is zero. This suggests that there are two methods for calculating *Core'* from B_1 and the *Etu(n)* values. In Approach 1,

$$Core' = -m_1 + B_1 \cdot \frac{1}{M}\sum_{n=1}^{M} Etu(n)$$

In Approach 2,

$$Core' = B_1 \cdot \frac{1}{M}\sum_{n=1}^{M} Etu(n)$$

2. Confidence Intervals for Correction Factor B_1

Now that we have B_1, m_1, and σ_1, we can calculate confidence intervals for the NR5 files. In Table 2, we have listed the 90% confidence intervals for these files, along with the corresponding values of *Core* and *Core'*. Note that the values of *Core'* fail to fall into the confidence interval only for files NR5-127-93, NR5123-93, and NR5-122-93. The fact that 3 out of 10 *Core'* values are outside the 90 % confidence intervals may seem troubling. However, the three files in question have more noise for some reason.

Table 2. Approach 2, 90 % Confidence Intervals for NR5 Files

NR5 File	Confidence Interval	Core	Core'
38-90	Conf(Core-.904 < Core' < Core+.608) = 90%	5.000	5.149
41-90	Conf(Core-.904 < Core' < Core+.608) = 90%	5.200	4.908
45-91	Conf(Core-.184 < Core' < Core+1.54) = 90%	4.500	5.230
56-91	Conf(Core-.112 < Core' < Core+.828) = 90%	5.500	6.034
66-91	Conf(Core-.636 < Core' < Core+.340) = 90%	5.300	4.894
110-92	Conf(Core-.838 < Core' < Core+.542) = 90%	4.800	4.291
122-93	Conf(Core-1.34 < Core' < Core+1.05) = 90%	4.400	5.632
123-93	Conf(Core-.838 < Core' < Core+.542) = 90%	5.100	3.754
127-93	Conf(Core-.838 < Core' < Core+.542) = 90%	3.800	4.760
128-93	Conf(Core-.993 < Core' < Core+.697) = 90%	5.200	4.525

C. Calculating *Core'* Using A Correction Factor, With Zero Bias

Approaches 1 and 2 have problems. Approach 1 always gives a positive value for $Core_1$, even when the *Etu(n)* values are zero. This may be unacceptable to Geophysics International customers. Approach 2, the straight average of *Etu(n}* multiplied by B_1, is biased. The values of *Core'* are more likely to be less than *Core* than greater than *Core*.

If we start out with the requirement that *Core'* is unbiased, which means that $E[\, e_1(n) \,] = 0$, we can derive **Approach 3**.

$$E[\ e_1(n)\ ] = 0,$$

$$E[\ B_2 \cdot Etu(n) - Core(n)\ ] = 0,$$

$$B_2 \sum_{n=1}^{N} Etu(n) = \sum_{n=1}^{N} Core(n),$$

$$B_2 = \frac{\sum_{n=1}^{N} Core(n)}{\sum_{n=1}^{N} Etu(n)}$$

In **Approach 3**, then

$$Core' = B_2 \cdot \frac{1}{M} \sum_{n=1}^{M} Etu(n)$$

Using the NR5 values again, we have calculated B_2, σ_2^2 and σ_2 as

$$B_2 = .8295,$$

$$\sigma_2^2 = 1.100,$$

$$\sigma_2 = 1.049$$

In Table 3, we have listed the 90 % confidence intervals for the NR5 files using these parameter values, along with the corresponding values of *Core* and *Core'*. Again, the values of *Core'* fail to fall into the confidence interval for files NR5-127-93, NR5-123-93, and NR5-122-93. In Approach 3, we see that the confidence interval is centered exactly at *Core'*. Also, the interval is wider and the estimates are worse than in Approach 2.

Table 3. Approach 3, 90 % Confidence Intervals for NR5 Files

NR5 File	Confidence Interval	Core	Core'
38-90	Conf(Core-.772 < Core' < Core+ .772) = 90%	5.000	5.309
41-90	Conf(Core-.772 < Core' < Core+ .772) = 90%	5.200	5.060
45-91	Conf(Core-1.73 < Core' < Core+ 1.73) = 90%	4.500	5.392
56-91	Conf(Core-.996 < Core' < Core+ .996) = 90%	5.500	6.221
66-91	Conf(Core-.498 < Core' < Core+ .498) = 90%	5.300	5.046
110-92	Conf(Core-.704 < Core' < Core+ .704) = 90%	4.800	4.424
122-93	Conf(Core-1.22 < Core' < Core+ 1.22) = 90%	4.400	5.806
123-93	Conf(Core-.704 < Core' < Core+ .704) = 90%	5.100	3 .871
127-93	Conf(Core-.704 < Core' < Core+ .704) = 90%	3.800	4.908
128-93	Conf(Core-.863 < Core' < Core+ .863) = 90%	5.200	4.666

E. Confidence Interval Subroutine

```
      subroutine  Confid(gamma,stdl,M,bias,Cl,C2)
      real x(8),cc(8),k
c
c     Assume that Core' is to be calculated from M estimates,
c     Core(n), each with a standard deviation of stdl and a bias
c     of bias.
c     Given gamma, std1, M, and the bias for Core', calculate
c     C1 and C2 such that Conf(Core-C1 < Core' < Core+C2) = gamma % get
C     constant c
C
C     gamma must be between 90. and 100.
C
      P=gamma/100.
      xCI) =.9
      cc(l)= 1.645
      x(2) = .925
      cc(2) = 1.78
      x(3)=.95
      ccC3)= 1.96
      x(4)=.975
      cc(4)=2.24
      x(5)= .99
      cc(5)=2.576
      x(6)= .999
      cc(6) =3.291
      do1i= 1,5
      if(x(i).le.P.and.p.le.x(i + l))then
      c = ((P-x(i))*cc(i + 1) + (x(i + l)-P)*cc(i))/(x(i + l)-x(i))
      k=(stdl/sqrt(l. *M)*c
      Cl =-bias-k
      C2=-bias+ k
      return
      endi f
1     continue
      return
      end
```