Semblance with Noise: The Statistic

by: Adam Gogacz (d-Analytics)

(1) GENERAL SETUP

The semblance statistic for a collection of n traces (time-series) $\{G_1, ..., G_n\}$ is defined as (https://wiki.seg.org/wiki/Semblance),

$$\mathcal{S}[j_0] := \frac{\sum_{j=j_0}^{j_0+m} \left(\sum_{i=1}^{n} G_i[t_{ij}] \right)^2}{n \sum_{j=j_0}^{j_0+m} \left(\sum_{i=1}^{n} G_i[t_{ij}]^2 \right)}$$

where by Cauchy–Schwarz inequality $0 \leq \mathcal{S}[j] \leq 1$. After some rearrangment and some (temporarily) simplifying notation,

$$\mathcal{S}[j_0] = \frac{\frac{1}{n^2} \sum_{j=j_0}^{j_0+m} \left(\sum_{i=1}^{n} G_i[t_{ij}] \right)^2}{\frac{1}{n^2}\ n \sum_{j=j_0}^{j_0+m} \left(\sum_{i=1}^{n} G_i[t_{ij}]^2 \right)} = \frac{\sum_{j=j_0}^{j_0+m} M_1[j]^2}{\sum_{j=j_0}^{j_0+m} M_2[j]}$$

where $G_i[t_{ij}]$ denotes an amplitude measurement by the i-th data recorder and at t_{ij}-th time increment

$$M_1[j] := \frac{1}{n} \sum_{i=1}^{n} G_i[t_{ij}] \quad \text{(sample mean/first moment)}$$

$$M_2[j] := \frac{1}{n} \sum_{i=1}^{n} G_i[t_{ij}]^2 \quad \text{(sample raw second moment)}$$

$$\hat{\sigma}^2[j] := \frac{1}{n} \sum_{i=1}^{n} \left(G_i[t_{ij}] - M_1[j] \right)^2 = M_2[j] - M_1[j]^2 \quad \text{(biased sample variance)}$$

and we end up with,

$$\mathcal{S}[j_0] = \frac{\sum_{j=j_0}^{j_0+m} \left(M_2[j] - \hat{\sigma}^2[j] \right)}{\sum_{j=j_0}^{j_0+m} M_2[j]} = 1 - \frac{\sum_{j=j_0}^{j_0+m} \hat{\sigma}^2[j]}{\sum_{j=j_0}^{j_0+m} M_2[j]}$$

Assuming that we carried out normal move-out (NMO) correction/alignment and all the traces are *imperfectly aligned* (also assuming insignificant amplitude variation along trace dimension), then,

$$G_i[t_{ij}] = a_{t_j} s_{t_j} + \varepsilon_{i\ t_j} + \gamma_{i\ t_j}$$

where

s_t :	transmitted signal
a_t :	decreasing function capturing all depth-dependent amplitude decay processes
ε_{it} :	independent zero-mean Gaussian random variables (background noise) with σ^2 variance
γ_{it} :	random variables due to NMO misalignment (independent of each other and of the background noise) following some unknown distribution $\mathcal{F}_\gamma$

so that,

$$\begin{aligned}
&\varepsilon_{i\ t_j} \overset{i.i.d.}{\sim} \mathcal{N}\left(0,\ \sigma_\varepsilon^2\right) && \Longrightarrow \mathbb{E}\left[\varepsilon_{i\ t_j}^2\right] = \sigma_\varepsilon^2 \quad \forall i, j \\
&\gamma_{i\ t_j} \overset{i.i.d.}{\sim} \mathcal{F}_\gamma \ \text{ s.t. } \ \mathbb{E}\left[\gamma_{i\ t_j}\right] = \mu_j && \Longrightarrow \mathbb{E}\left[\gamma_{i\ t_j}^2\right] = \mathrm{Var}\left[\gamma_{i\ t_j}\right] = \sigma_{\gamma_j}^2 \quad \forall i, j
\end{aligned}$$

then,

$$\begin{aligned}
M_1[j] &= \frac{1}{n}\sum_{i=1}^{n}\left(a_{t_j}s_{t_j} + \varepsilon_{i\ t_j} + \gamma_{i\ t_j}\right) = a_{t_j}s_{t_j} + \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i\ t_j} + \frac{1}{n}\sum_{i=1}^{n}\gamma_{i\ t_j} \\
\hat{\sigma}^2[j] &= \frac{1}{n}\sum_{i=1}^{n}\left(G_i[t_{ij}] - M_1[j]\right)^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\varepsilon_{i\ t_j} - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i\ t_j} + \gamma_{i\ t_j} - \frac{1}{n}\sum_{i=1}^{n}\gamma_{i\ t_j}\right)^2
\end{aligned}$$

Therefore, the semblence statistic for NMO corrected (aligned) traces is,

$$\mathcal{S}[j_0] = 1 - \frac{\displaystyle\sum_{j=j_0}^{j_0+m}\sum_{i=1}^{n}\left(\varepsilon_{i\ t_j} - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i\ t_j} + \gamma_{i\ t_j} - \frac{1}{n}\sum_{i=1}^{n}\gamma_{i\ t_j}\right)^2}{\displaystyle\sum_{j=j_0}^{j_0+m}\sum_{i=1}^{n}\left(a_{t_j}s_{t_j} + \varepsilon_{i\ t_j} + \gamma_{i\ t_j}\right)^2}$$

For n large enough the Law of Large Numbers (LLN) provides us with the approximations,

$$
\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i\ t_j} \approx \mathbb{E}\left[\varepsilon_{i\ t_j}\right] = 0 \quad \forall\ i,j \\
&\frac{1}{n}\sum_{i=1}^{n}\varepsilon_{i\ t_j}^2 \approx \mathbb{E}\left[\varepsilon_{i\ t_j}^2\right] = \sigma_{\varepsilon} \quad \forall\ i,j \\
&\frac{1}{n}\sum_{i=1}^{n}\gamma_{i\ t_j} \approx \mathbb{E}\left[\gamma_{i\ t_j}\right] = \mu_j \quad \forall\ i,j \\
&\frac{1}{n}\sum_{i=1}^{n}(\left(\gamma_{i\ t_j} - \frac{1}{n}\sum_{i=1}^{n}\gamma_{i\ t_j}\right)^2 \approx \mathbb{E}\left[\left(\gamma_{i\ t_j} - \mathbb{E}[\gamma_{i\ t_j}]\right)^2\right] = \sigma_{\gamma_j} \quad \forall\ i,j
\end{aligned}
$$

Hence, for n sufficiently large LLN gives,

$$
\boxed{\mathcal{S}[j_0] \approx 1 - \frac{\sum_{j=j_0}^{j_0+m}\mathbb{E}\left[\left(\varepsilon_{i\ t_j} + \gamma_{i\ t_j} - \mu_j\right)^2\right]}{\sum_{j=j_0}^{j_0+m}\mathbb{E}\left[\left(a_{t_j}s_{t_j} + \varepsilon_{i\ t_j} + \gamma_{i\ t_j}\right)^2\right]}}
$$

(2) SPECIAL CASE: $m = 0$ and no background noise ($\varepsilon_{i\ t_j} = 0$)

In a (background) noise-free and imperfectly NMO corrected traces,

$$
\begin{aligned}
\varepsilon_{it_j} &= 0 \\
\gamma_{i\ t_j} &\neq 0
\end{aligned}
$$

and n large enough, the boxed equation reduces to,

$$
\mathcal{S}[j] \approx 1 - \frac{\text{Var}\left[\gamma_{i\ t_j}\right]}{\mathbb{E}\left[\left(a_{t_j}s_{t_j} + \gamma_{i\ t_j}\right)^2\right]} = 1 - \frac{\sigma_{\gamma_j}^2}{a_{t_j}^2 s_{t_j}^2 + \sigma_{\gamma_j}^2}
$$

The above equation shows that for $s_{t_j} \neq 0$ the semblence statistic is *maximized* whenever the perturbations in the alignment (variance of the perturbation) tend to zero; that is, whenever $\sigma_{\gamma_j}^2 \to 0$. Therefore, in low or no noise condition the semblance statistic is maximized whenever the tested velocity model converges to the true model (perfectly NMO corrected traces).

(3) SPECIAL CASE: $m = 0$ and perfect NMO alignment ($\gamma_{i\ t_j} = 0$)

For perfectly NMO corrected traces,

$$
\begin{aligned}
\mathbb{E}\left[\gamma_{i\ t_j}\right] &= \mu_j = 0 \\
\text{Var}\left[\gamma_{i\ t_j}\right] &= 0
\end{aligned}
$$

the boxed equation reduces to,

$$\mathcal{S}[j] \approx 1 - \frac{\mathbb{E}\left[\left(\varepsilon_{i\ t_j}\right)^2\right]}{\mathbb{E}\left[\left(a_{t_j} s_{t_j} + \varepsilon_{i\ t_j}\right)^2\right]} = 1 - \frac{\sigma_\varepsilon^2}{a_{t_j}^2 s_{t_j}^2 + \sigma_\varepsilon^2}$$

The above equation shows that for n large enough the semblence statistic is always less than 1 and decays towards 0 as $a_{t_j} s_{t_j} \to 0$. Moreover, as the signal-to-background noise ratio decreases, the performance of the semblance statistic diminishes.

(4) DISCUSSION

The above analysis shows that where signal-to-noise ratio (S/N) is low, the semblance statistic performs quite poorly. Note that low S/N might occur in two different ways, 1. scattering from weak or deep objects (i.e. reflectors, spatialy confined anomalies) 2. wavelet is close to its zero-crossing (wavelets are assumed to be zero-mean)

Hence, semblance analysis is best carried out on peaks of the transmitted wavelet or the length of the time-gate (m) should be about $^1/_4$ of the dominant wavelength.