Electronic Appendices: to appear as supplementary material on journal archive

Appendix 1: Bayesian update formulas for approximate multi-Gaussian posteriors including the noise parameters

The standard linear model with a Jeffreys prior on a single noise parameter has a simple analytical posterior featuring a (right skewed) Inv–χ^2 marginal distribution for the noise (Gelman et al., 1995, section 8.9). Although this posterior is not strictly integrable to obtain a marginal model likelihood, a Gaussian approximation at the MAP point will be. When there are multiple noise parameters and some parts of the likelihood are not scaled by the unknown noise, the posterior is somewhat messier. The amount of data commonly available in well–tie problems is usually sufficient to make the posterior distribution of the noise components fairly compact, so will develop the relevant formulae on the assumption that a full multi–Gaussian form for the joint model and noise parameters is adequate.

Since the noise posteriors are usually right skewed, it is sensible to work in terms of the joint model vector $\mathbf{M} = \{\mathbf{m}, \mathbf{m}^{(\mathrm{n})}\}$ where $\mathbf{m}^{(\mathrm{n})}$ is defined as $\mathbf{m}_j^{(\mathrm{n})} = \log(\sigma_{\mathrm{stack}\ j})$, the noise level for stack j. The non–noise components $\mathbf{m}$ are formed from the wavelet components $\mathbf{a}_w$ (and coupling parameters for the coupled near/far wavelet mode of operation), time to depth map knots $\boldsymbol{\tau}$, positioning and registration errors $\boldsymbol{\Delta}\mathbf{r}_R$ and AVO scale term B respectively:

$$\mathbf{m} \equiv \{\mathbf{a}_w, \boldsymbol{\tau}, \boldsymbol{\Delta}\mathbf{r}_R, B\}.$$

We assert a Gaussian prior for $\mathbf{M}$ of form $P(\mathbf{M}) = N(\bar{\mathbf{M}}, C_M)$, where $C_M = \mathrm{diag}\{C_p, C_p^{(n)}\}$ and the $C_p^{(n)}$ entries for noise components $\mathbf{m}^{(\mathrm{n})}$ are very weak (we choose the mean to be about the log of the signal RMS, and the standard deviation large on the log scale), and absorb (for computational convenience) the Jeffrey's noise prior and wavelet peak/phase prior terms into the collected likelihood

$$L(\mathbf{y}|M) = \frac{P_{\mathrm{Jeffreys}}(\mathbf{m}^{(\mathrm{n})})}{|C_D(\mathbf{m}^{(\mathrm{n})})|^{1/2}} \exp(-(\mathbf{y} - f(\mathbf{m}))^T C_D(\mathbf{m}^{(\mathrm{n})})^{-1}(\mathbf{y} - f(\mathbf{m}))/2).$$

This likelihood collects all non–Gaussian terms in (7), so $\mathbf{y}$ is an 'equivalent data' vector, formed by the concatenation of the seismic data for all wells and stacks, sonic–log derived interval velocities, and (from the prior) desired phase/timing characteristics of the wavelet. Similarly $f(\mathbf{m})$ is the concatenation of the synthetic seismics, model–derived interval velocities, and the wavelet–phase/timing measures. Only that part of $C_D(\mathbf{m}^{(\mathrm{n})})$ pertaining to the seismic mismatch depends on $\mathbf{m}^{(\mathrm{n})}$. Note that C_D is diagonal, with $\exp(2\mathbf{m}_j^{(\mathrm{n})})$ terms for the seismic mismatch on stack j, σ^2_{Vint} terms for the interval velocities, and $\sigma^2_{\mathrm{phase}}, \sigma^2_{\mathrm{peak\text{-}arrival}}$ terms for the wavelet phase/timing.

Under the logarithmic change of variables, the Jacobean and coefficient of the above equation conspire to yield the log-likelihood

$$-\log(L(\mathbf{y}|M)) = \mathbf{e}^T.\mathbf{m}^{(\mathrm{n})} + \frac{1}{2}(\mathbf{y} - f(\mathbf{m}))^T C_D(\mathbf{m}^{(n)})^{-1}(\mathbf{y} - f(m)).$$

where $e_j = N_{\mathrm{stack}\ j}$, the number of seismic mismatch data points for stack j. If all terms in the log–likelihood and prior are expanded to quadratic order in $\{\mathbf{m}, \mathbf{m}^{(\mathrm{n})}\}$ around some current guess $\mathbf{M}_0 = \{\mathbf{m}_0, \mathbf{m}_0^{(\mathrm{n})}\}$ and collected, the resulting quadratic form corresponds to a posterior Gaussian distribution for $\mathbf{M}$ which may be solved using the following expressions:

$$\Delta\tilde{y}_i = \frac{\mathbf{y}_i - f_i(\mathbf{m}_0)}{C_{D,ii}^{1/2}} \tag{9}$$

$$\tilde{X}_{ij} = \frac{f_i(\mathbf{m}_0 + \delta m_j) - f_i(\mathbf{m}_0)}{\delta m_j C_{D,ii}^{1/2}} \quad \text{(sensitivity matrix from finite diffs.)} \tag{10}$$

$$\hat{Q}_{i,j} = -2\delta_{I(i),j}\Delta\tilde{\mathbf{y}}_j \qquad (N_{\mathrm{stacks}}\times \dim(\mathbf{y}) \text{ matrix}) \tag{11}$$

$$\tilde{P}_{ij} = 2\sum_k (\Delta\tilde{\mathbf{y}}_k)^2 \delta_{I(k),i}\delta_{I(k),j} \qquad (N_{\mathrm{stacks}} \times N_{\mathrm{stacks}} \text{ matrix}) \tag{12}$$

where $I(i)$ is an indicator variable that is the index of the stack–noise parameter in $\mathbf{m}^{(\mathrm{n})}$ that applies to component i in the 'data' vector $\mathbf{y}$. It is zero for non seismic–mismatch components.

With these definitions, the linear system for the model update $\Delta\mathbf{M}$ is

$$\begin{pmatrix} C_p^{-1} + \tilde{X}^T\tilde{X} & -\tilde{X}^T\hat{Q}^T \\ -\hat{Q}\tilde{X} & C_p^{(\mathrm{n})} + \tilde{P} \end{pmatrix} \cdot \begin{pmatrix} \Delta\mathbf{m} \\ \Delta\mathbf{m}^{(n)} \end{pmatrix} = \begin{pmatrix} C_P^{-1}(\bar{\mathbf{m}} - \mathbf{m}_0) + \tilde{X}^T\Delta\tilde{y} \\ -\mathbf{e} - C_p^{(n)-1}(\mathbf{m}_0^{(\mathrm{n})} - \bar{\mathbf{m}}^{(\mathrm{n})}) - \frac{1}{2}\hat{Q}\Delta\tilde{y} \end{pmatrix}$$

(13)

The update then has mean $\bar{M}' = \mathbf{M}_0 + \Delta\mathbf{M}$ and a covariance $\tilde{C}$ which is the inverse of the coefficient matrix in (13):

$$\tilde{C} \equiv \begin{pmatrix} C_p^{-1} + \tilde{X}^T\tilde{X} & -\tilde{X}^T\hat{Q}^T \\ -\hat{Q}\tilde{X} & C_p^{(\mathrm{n})} + \tilde{P} \end{pmatrix}^{-1}. \quad (14)$$

These update formula form the basis of a Newton scheme for finding the maximum aposteriori point for a particular model with wavelet vector $\mathbf{a}_w$. The RHS of (13) is effectively a gradient of the posterior density, so the updates vanish when the mode is reached. The actual implementation is a typical Gauss–Newton scheme with line searches in the Newton direction and additional step-length limits applied for big Newton steps. Laplace estimates based on the determinate of the final $\tilde{C}$ are used in forming the marginal model likelihood.

Appendix 2: Data input forms

The data needed to perform the wavelet extraction should be in the following forms

- Configuration file. For a particular extraction, the configuration is defined by an XML file which sets up all necessary parameter choices and file names. It is controlled by a schema file (XSD), so configuration is simple with the BHP XML schema-driven editor supplied with the distribution. The schema is located at `au/csiro/JamesGunning/waveletextraction/xsd/` relative to the root of the distribution, and the editor at `scripts/xmledit.sh`. Certain self–documenting (SU style) runtime flags are able to modify the behavior at execution for frequently chosen options. We recommend users read the self docs carefully and follow the examples.
- Seismic data as a 'mini–cube', for each well, in big–endian SU format. Inline, crossline and gx,gy (group coordinate) header fields need to be correctly set. The cube should be big enough to safely enclose each well on the desired extraction time interval, but not so large as to swamp memory (10x10 cubes are typically fine for straight holes). If x,y headers are not available, a triplet

of non–collinear points can be specified in the XML file which enables the code to map inline,crossline to x,y. Single line or trace seismic files are also acceptable, provided suitable flags are set in the XML.

- Log data is read in ascii, either standard ascii LAS files or the geoEAS format used by GSLIB (Deutsch and Journal, 1998). The critical depth to be specified in these files is measured depth (MD) along the well trajectory. There must be density and DT (slowness) fields for the sonic p–wave and shear logs. A rudimentary commandline switch (`--fake-Vs`) will concoct a placeholder shear log from the p–wave log if shear is not available, which is useful for near–normal incidence extractions.
- UTM files with the well trajectories. These are geoEAS files, which must have x, y, MD and TVD columns. Kelly bushing data will be required to convert to true depth TVD.
- Checkshot data. Again, geoEAS format. We require MD, T, sigma_T columns.
- Markers (optional): geoEAS format also, with MD, T, sigma_T columns.

A note on units: the code performs automatic time units detection for time data, which is expected to be input in either s or ms. All density (mass) units cancel internally. Equation (2) involves a velocity ratio from log data to stacking–velocity V_{st} (see eqn (3)), and the code will again attempt to auto–detect a feet vs meters difference here (logs are commonly in μS/ft, V_{st} may be in m/s). Another velocity ratio occurs in (3.1.2); again, auto–detection is performed, since $\mathbf{V}_{\mathrm{int}}(\boldsymbol{\tau})$ is derived from a checkshot that may be imperial or metric. The auto–detection code is based on the assumption fact that velocities will be specified in m/s or ft/s, and that sonic logs will be in μS/ft or (very unusually) metric.

Users are recommended to look at the examples in the distribution closely when setting up their own problems.

Appendix 3: Outputs

A variety of outputs are generated by the code. Many of these are selectable by commandline switches (the code self–documents like SU codes). All SU outputs are big–endian. In roughly diminishing order of importance, these are

- ASCII dumps of the cross-registered logs, observed and synthetic seismic, for each well and wavelet span (typically `MLgraphing-file.*`). Used as direct input for the ExtractionViewer module.

- Maximum likelihood (ML) wavelet, SU format, for each wavelet span, and overall (`MLwavelet.*.su`).
- ASCII files of the most likely parameter values and uncertainties (diagonal entries of the posterior covariance matrix), for each wavelet span, and overall (`MLparameters.*.txt`).
- Multiple realizations of the wavelets from the posterior, also SU format (`wavelet_realisations.su`).
- Maximum likelihood 5–trace seismic sets along the well trajectory, for the five quantities {observed–seismic, synthetic seismic, reflectivity, x, y}, again, for each wavelet span, and overall (`MLsynth_plus_true_seis.*.su`) . These are very useful for third party software applications.
- ASCII files for the most likely time-to–depth map, for each well, for each wavelet span, and overall (`MLtime-to-depth-map.*.txt`)
- ASCII files of the most likely wavelet phase spectrum, for each wavelet span, and overall (`MLwavelet-phase-spectrum*`).
- ASCII dumps of the well log blocking (`*_blocked.txt`).
- ASCII dumps of negative log–posterior cross-sections of the posterior surface, shown as slices for each model parameter (`posteriorXsection*.txt`). These are useful for checking for possible multimodality, or serious non–linearity.

Appendix 4: Issues in positional uncertainty modeling

When using the extraction model with general positioning/registration parameters, this amounts to using a generalized non–linear regression of form

$$\mathbf{S}_{\mathrm{obs}}(m) = \mathbf{r}(m) * \mathbf{w}(m) + \mathbf{e_n},$$

where some parameters in m determine (by continuous interpolation) the local trace data $\mathbf{S}_{\mathrm{obs}}(m)$ to be used in the wavelet extraction.

In this kind of model, one has to be wary of the optimization 'latching' on to a particularly favorable location parameter which is controlled merely by chance rather than a genuinely better fit to the regression model. A better insight into this problem may be obtained by examining a simplified discrete 'choose–the–data–set' problem, in which spatial fluctuations in y are mimicked by a choice of one of two data sets y_1 or y_2, and the 'choice' parameter is controlled by the Bayes factors at the maximum likelihood point. We can then examine how the expected fluctuations in the Bayes factor affect the uncertainty in the 'choice' parameter by Monte–Carlo methods.

A canonical linear regression model for a specific data set y would be

$$y = X.m + \mathbf{e_n}$$

where X is the design matrix and $e_n \sim N(0, \sigma^2)$ is the noise. The standard likelihood is

$$-\log(L(d|m)) \sim (y - X.m)^T (y - X.m)/2\sigma^2 + n\log(\sigma)$$

A Jeffrey's prior for the noise variance σ^2, coupled with an independent Gaussian $N(0, \Sigma_m)$ for the coefficients leads to an improper posterior marginal probability for the model (Gelman et al., 1995), so we use a class of conjugate priors – Zellners' g-priors (Zellner, 1986) – which leads to simple analytic forms for the posterior. These priors take the form

$$p(\sigma^2) \sim 1/\sigma^2 \tag{15}$$
$$p(m|\sigma^2) = N(0, g\sigma^2(X^T X)^{-1}) \tag{16}$$

where g is a single hyper-parameter. For the linear model, the posterior marginal probability has the simple form

$$\Pi = \int (L(d|m)p(m|\sigma^2)p(\sigma^2)dmd\sigma^2 \sim (1+g)^{(n-k-1)/2}(1+g(1-R^2))^{-(n-1)/2}$$

where there are n data points, k parameters in m, and R^2 is the usual coefficient of determination. The Bayes factor for two identical models fitted to different data sets (traces) y_1 and y_2 is then

$$B_{12} = \left(\frac{(1+g(1-R_1^2)}{(1+g(1-R_2^2))}\right)^{-(n-1)/2}.$$

It is instructive to evaluate this quantity for realistic models of the relative fluctuation in R_1 and R_2 as we imagine extracting the data sets y_1, y_2 from a trace in a typical near–well region. To illustrate, we have fitted the 'noisy straight line' data sets(s)

$$y_{(1,2),i} = 1 + i/n + q_i + \epsilon_{(1,2),i} \qquad i = 1, 2 \ldots n$$

for $n = 100$ data points $i = 1, 2 \ldots n$, with X describing a standard linear model, $q \sim N(0, 0.25^2)$ a *fixed* (shared) error realization of relative RMS amplitude typical for a well–tie, and the two samples y_1, y_2 distinguished by the $\epsilon_{(1,2),i}$ term, with the setting $\epsilon_{(1,2),i} \sim N(0, 0.05^2)$ emulating small spatial fluctuations.

Different samples of y are generated by different realisations of ϵ. We have used the recommended setting $g = n$ to set the prior covariance $g(X^T X)^{-1}$ to be $O(1)$ (see Zellner (1986) and Clyde [4] for a more extensive discussion of the role of g). Under this model, the Monte–Carlo distribution of B_{12} formed when y_1 and y_2 are drawn from different realisations of ϵ (for a fixed, random q), is illustrated in Fig. 6. Large fluctuations in the Bayes factor associated with 'position' (of which ϵ is the surrogate) are induced by the relatively weak (5%) additional noise. It is easy to imagine that the deep local minima associated with these large Bayes factors will also occur in the full nonlinear wavelet extraction problem, where the 'choice' parameter is a continuous spatial variables, rather than a discrete label. If so, it is likely that strongly constrained positioning parameters (associated with deep local minima in the map of the Bayes factor as we move around in the near–well region) are not trustworthy.

[Fig. 6 about here.]

In summary, we recommend using the lateral position modeling capabilities of this package with extreme caution. Particular skepticism ought to be exercised with respect to the MAP Hessian–derived standard deviations attached to the positioning parameters.

End of electronic appendices

[4] Clyde, M., George, E. I., 2003. Model uncertainty. Statistical and Applied Mathematical Sciences Institute Technical Report #2003–16, see `www.samsi.info`.

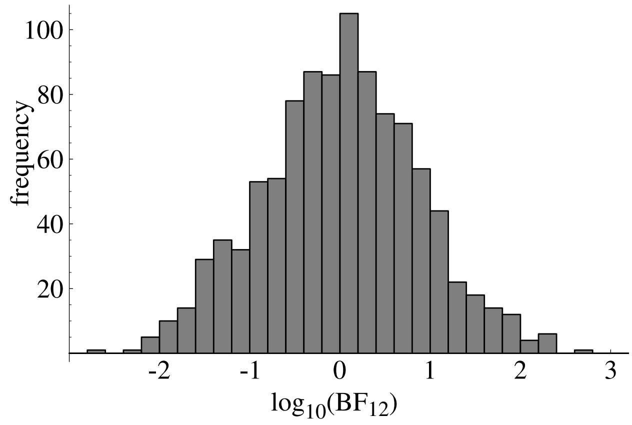

Fig. 6. Variation in Bayes factor due to trace variations of the order of 5% added to a linear model. Note the logarithmic scale. The wide variations in the Bayes factor would lead to unjustifiably strong conclusions about the optimal positioning of the seismic data if their (at least partial) origin in pure chance is not considered.