Delivery–Extractor: A new open-source wavelet extraction and well tie program

James Gunning, CSIRO Petroleum, and Michael Glinsky, BHP Billiton

Summary

We present a new *open–source* Bayesian wavelet extraction program for deriving wavelets from seismic and well–log information. This code is designed to complement the open–source model–based Bayesian inversion code *Delivery* (Gunning and Glinsky; Gunning, 2003), recently presented at the EAGE meeting in Stavanger. The *Delivery–Extractor* program approaches the well–tie problem from a Bayesian viewpoint, which naturally integrates prior knowledge about the well tie in the form of marker constraints, VSP data, phase prejudices, or plausible interval velocities. The code can perform simultaneous extractions at multiple (possibly deviated) wells, for multiple offsets (using a linearised Zoeppritz reflectivity), and can estimate additional uncertainty parameters such as time–registration errors for stacks or well–location errors caused by imaging problems. The code produces distribution details for the mis–tie, or noise, amplitude (critical for inversion studies), and can produce multiple realisations of the extracted wavelets from the Bayesian posterior, showing the uncertainty in the wavelet scaling and extent, the time–to–depth map, and the noise parameters for each stack.

Introduction

Seismic inversions are inescapably and critically dependent on the estimate of the source signature or wavelet. Though only the most naive characteristics of the reservoir geometry are robust to poor estimations of the wavelet shape, it is not unusual to see expensive and time–consuming inversion calculations performed using badly calibrated wavelets. In order to produce the best possible inversions from *Delivery*, we felt it was necessary to write a high quality Bayesian wavelet extraction code to produce optimal estimations of the wavelet for inversion studies and other possible applications, using many of the proven and successful ideas in the *Delivery* inverter.

A few remarks about our views on wavelet extraction and inversion are pertinent.

- From a modelling and computational viewpoint, fully probabilistic Bayesian inversion is possible only with a relatively simple forward model of wave propagation. For field scale studies, this restricts us to the traditional convolutional model, which amounts to modelling primary reflections only, or the "small contrast" approximation. It follows then that any forward modelling parameters (such as the wavelet) need only be accurate to the same order. Extracting wavelets using the full Zoeppritz equations for an inversion program using convolutional models is both inconsistent and pointless. We use a linearised Zoeppritz equation with terms to θ^2 in incidence angle
- Experience has consistently shown the well–tie problem to be 'data–poor', i.e. there are usually not many independent data points compared to the number of parameters we wish to estimate. Extractions must usually be confined to a window of about 0.5 seconds or less (to avoid modelling across significant dispersion), yielding usually O(100) data points or fewer, and the sum total of the parameters to be estimated is often many tens. This means that estimates of the wavelet shape are often very prone to be 'overfitted' by very flexible models. This makes the hypothesis and modelling of 'spatially–varying' wavelets between wells very statistically dubious, and highly vulnerable to overfitting. Transverse variations in wavelet character do occur, and for good reasons, but it is extremely difficult to model (and interpolate) such variations in ways that are statistically defendable. We feel it is safer to work with the hypothesis of spatial stationarity and possibly exaggerate the estimate of the noise strength, since this will prevent the driving of subsequent inversions into overfitting by a misleadingly aggressive S/N ratio.
- The length of the wavelet is never known a–priori, and it is desirable for this quantity (or its distribution) to be estimated using some canonical modern approach like Bayesian Model Selection (BMS) (Denison et al., 2002). In practice, this will involve computing well–ties for all possible wavelet lengths and ordering them using Bayes factors. The most likely wavelet from the BMS calculation will be a sensible compromise between over and under fitting.

Our approach to wavelet extraction has much in common with the work of Buland and Omre (2003). We incorporate the additional (and significant) uncertainty of the wavelet span, time registration errors, well positioning errors, AVO related tweaks to the reflectivity equation, and additional constraints on wavelet phase and interval–velocity errors.

Parameters of the problem

The basic problem in wavelet extraction is to find a wavelet whose time–convolution with the reflection coefficients computed from well log data yields a good match to the stacked migrated seismic data at the well location. This leads to a non–linear regression problem for a set of parameters **m** comprising (a) the coefficients describing the wavelet plus (b) any free parameters that contribute to the mapping of the well–log data in depth to the time axis. Usually the time–to–depth mapping (e.g. a piecewise linear curve) is roughly established from a VSP or set of stacking/migration velocities, and local 'stretch–or–squeeze' adjustments $\mathbf{\Delta t}$ are applied at the curve 'knot' points (checkshots or markers) to improve the well tie. A global time–offset parameter is often useful here too. Also (c) certain parameters describing uncertainty in the well placement in space or the registration of the seismic in time may be included. There are several constraints in the regression, including (1) the range of permissible adjustments to the time–depth map ought to be constrained by sonic log information ('plausible interval velocities'). (2) It may be desired to constrain the wavelet phase for reasons to do with the processing.

The Bayesian approach to this regression problem is to regard the 'best' wavelet as a subvector of the maximum aposteriori (MAP) point of the posterior distribution for $\mathbf{m}$, given the 'data' D:

$$\Pi(\mathbf{m}|D) = p(\mathbf{m}) L_{\text{mistie}}(\mathbf{m}|D) L_{\text{interval-velocities}}(\mathbf{m}|D) \times L_{\text{phase}}(\mathbf{m}). \quad (1)$$

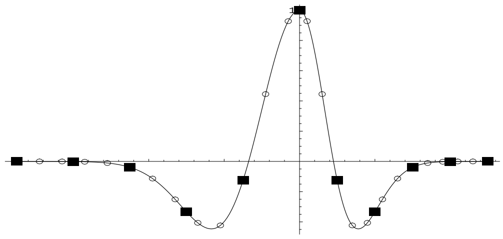
Here $p(\mathbf{m})$ is the prior distribution of the model parameters, $L_{\text{mistie}}(\mathbf{m}|D)$ is a likelihood function which expresses the likelihood of the model in terms of the well–tie seismic mismatch, $L_{\text{interval-velocities}}(\mathbf{m}|D)$ expresses the likelihood of the model in terms of the deviations of the interval velocities from comparable ones derived from the well log, and $L_{\text{phase}}(\mathbf{m})$ is a likelihood enforcing any phase constraints desired on the wavelet. We briefly describe each of these terms.

The wavelet $\mathbf{w}$ is parameterised in terms of a set of samples $\mathbf{m}_{w,i}$, $i = -M \ldots N$, spaced at the Nyquist rate associated with the seismic band edge (typically about $\delta t = 1/(4 f_{\text{peak}})$). The samples for the wavelet at the seismic data rate (e.g. 2,4 ms) are generated by cubic splines with zero-derivative endpoint conditions. See fig. 1. We use a generously wide Gaussian of mean zero for the Bayesian prior of $\mathbf{m}_{w,i}$. Note there are fewer fundamental parameters than seismic samples. This parameterisation enforces sensible bandwidth constraints and the necessary tapering. Cubic splines are a linear mapping, so the coefficients at the seismic sampling scale are related to the underlying coefficients $\mathbf{m}_w$ linearly, and the two indices M, N then define an set of wavelet models.

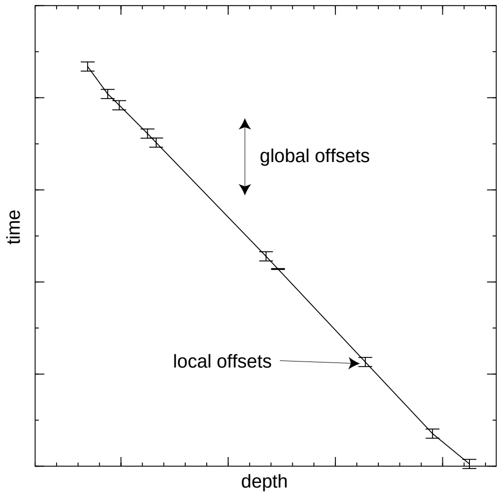
The stretch–and–squeeze parameters are modelled as having independent Gaussian priors $N(0, \sigma_{ij})$, for knot point j on well i, and typically set to a few ms. A global shift of the well registration can be modelled, as well as transverse shifts in space. These parameters have reasonable zero–mean Gaussian priors too.

The $L_{\text{mistie}}(\mathbf{m}|D)$ likelihood for the seismic mistie has the form

$$-2\log(L_{\text{mistie}}) \sim \sum_{\text{wells i, stacks k}} \left\{ \frac{(\mathbf{S}_{ik} - \mathbf{r}_{\mathbf{i}}(\mathbf{m}) * \mathbf{w}(\mathbf{m}))^2}{\sigma_k^2} + (n_i + 1)\log(\sigma_k) \right\}$$



a) Wavelet parameterization



b) Time to depth map with stretch-and-squeeze points

Fig. 1: (a) Parameterisation of wavelet in term of coefficients at Nyquist spacing associated with band edge (black boxes), and resulting coefficients generated at the seismic time-sampling rate (circles) by cubic interpolation. (b) Parameterisation of the time to depth map using stretch–and–squeeze points, usually placed at geological markers.

where S is the seismic data and n is the number of mistie evaluations in the well–tie interval. The reflectivity $\mathbf{r}$ is computed from automatically–blocked density and sonic logs using Backus averaging and the linearised Zoeppritz equation appropriate for the stack angle of each stack.

Tweaks in the stretch–and–squeeze parameters will change the effective interval velocities in the time to depth map. The $L_{\text{interval-velocities}}(\mathbf{m}|D)$ likelihood is a Gaussian model that enforces the need for these interval velocities to be within some reasonable (say 5% std. deviation) of the upscaled sonic log interval velocities. Similarly, the phase–constraint likelihood $L_{\text{phase}}(\mathbf{m})$ can enforce common requirements such as a constant, or zero phase wavelet, by suitable penalties on the phase spectrum corresponding to the wavelet embedded in $\mathbf{m}$.

Since the forward model is a trivial convolution, the MAP point can be found a using standard quasi–Newton optimiser and the uncertainty in all the parameters is found by computing the covariance of a Gaussian approximation to the posterior at this point. This almost always provides a very good approximation to the posterior distribution of the parameters, for a given wavelet span. The Bayes factor for each possible wavelet span is computed using the Laplace approximation (Raftery, 1996). This then produces the marginal probability of each wavelet span, which can be used to generate stochastic samples of the wavelet span on top of the intrinsic wavelet uncertainty.

The code will produce MAP estimates of the wavelet and other parameters, MAP well–synthetics, interval velocities, error–bars, posterior cross–sections and a host of diagnostics. Stochastic realisations of all the parameters can be generated.

Tests

The code has been tested successfully on a number of simple synthetic test cases to check the wavelet–span detection algorithms. It has been applied to several field cases with fixed time–to–depth maps and successfully reproduced wavelets from two standard commercial packages. The additional freedom in the time–to–depth maps has led to significantly better ties that are still guaranteed to remain consistent with log data.

A simple example is illustrated in figure 2, which shows the density, and sonic logs registered against the seismic and synthetic. An additional trace shows the effective sonic log corresponding to the interval velocities, with the marker points clearly evident. This is one pane of a multiwell tie, the other corresponding to a sidetrack from the same well location.

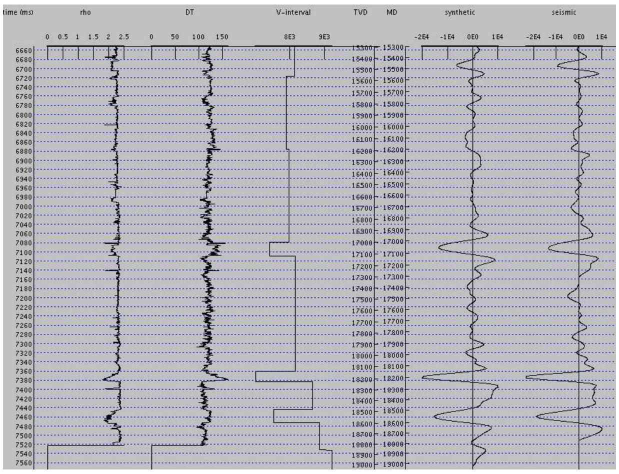


Fig. 2: Typical well–tie diagnostic output

The extracted wavelet from this tie is significantly shorter than that obtained from the commercial packages (see figure 3). Experience has shown this to be generic behaviour: much of the precursor and coda energy in standard wavelet derivations is not statistically significant, and can be legitimately truncated.

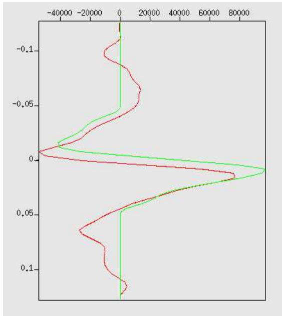


Fig. 3: Extracted wavelet (green) compared to standard commercial extraction (red). The wavelets found using Bayesian model selection are generally much shorter.

The code

The code is written in java, and uses highly efficient public domain libraries for the computationally heavy FFT, optimization, and linear algebra work. It is platform independent and driven by a simple XML file, and a simple GUI is provided for visualizing the maximum likelihood output. Outputs are compatible with Seismic Unix, *Delivery*, and the INT viewer.

The code comes bundled with the *Delivery* inverter, form CSIRO Petroleum's main website (Gunning, 2003).

Conclusions

- This wavelet extraction code is a major contribution to the repertoire of open–source tools for the geophysical community. It is capable of multi–stack, multiple–well and deviated well wavelet derivations. It features a fully integrated Bayesian approach to the coupled uncertainties in wavelet estimation, the time–to–depth map, vertical ans transverse registration errors, and noise estimation.

- Workers involved in inversion studies should have a particular interest in this code, as it provides maximum likelihood estimates of the both the wavelet and the noise level. Optimum estimates of the precursor and coda times are also invaluable for the purposes of building reservoir–centered layer–based models such as those used by *Delivery*.

- The community is welcome to download the code (Gunning, 2003) and trial it. Suggestions, bug reports and contributed improvements within the scope of the documented open–source agreement are welcome.

Acknowledgements

James Gunning would like to thank the BHP–Billiton quantitative interpretation project for funding.

References

Buland, A. and Omre, H. Bayesian wavelet estimation from seismic and well data. *Geophysics*, 68(6):2000–2009, 2003.

Denison, D. G. T. et al. *Bayesian Methods for Nonlinear Classification and Regression*. Wiley, 2002.

Gunning, J. **Delivery** website: follow links from `http://www.petroleum.csiro.au`. 2003.

Gunning, J. and Glinsky, M. **Delivery**: an open–source model–based Bayesian seismic inversion program. 'Computers and Geosciences', in press.

Raftery, A. E. Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice*. Chapman and Hall, 1996.