Wavelet Extraction: an essay in model selection and uncertainty

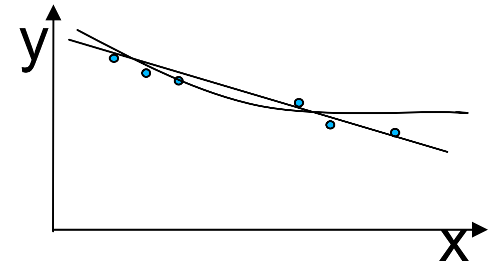
James Gunning, CSIRO Petroleum, Australia

Outline

- Model selection, inference and prediction: Generalities
- Wavelet extraction
  - Model formulation
  - Generic inversion/inference features for single models
  - Model choice scenarios. 3 examples
- Morals & conclusions

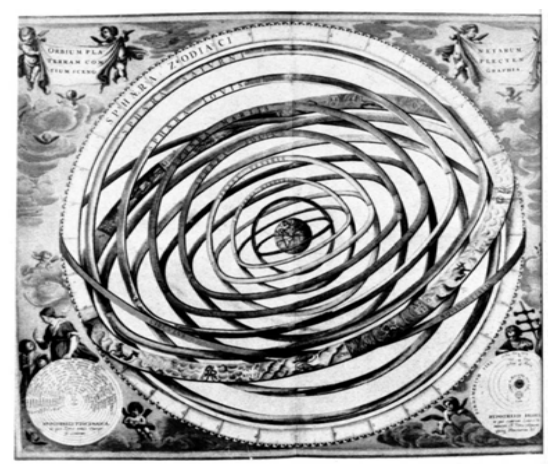
Model selection

- Classic statistics problem
  - line ? or parabola? or sextic?



- Foundations of science
  - Newton or Ptolemy?

F = m **a**



Generalities

- Geostatistics is only one branch of multivariate statistics.
  - Every result of significance can be derived from conventional multivariate estimation theory. A Bayesian spin is helpful.
  - The rest of the world understands us better if we write things this way.
- As statisticians, we have the duties of
  - Model inference
  - Significance testing
  - Model refinement/testing
  - Making predictions with error estimates

Model selection & Bayes factors

- Suppose there are $k=1\ldots N$ models to explain data D, with parameters m_k
- The model has prior $P(m_k|k)P(k)$, likelihood $L(D|m_k,k)$
- Here's *Three* very useful things:
  - Marginal model likelihood:

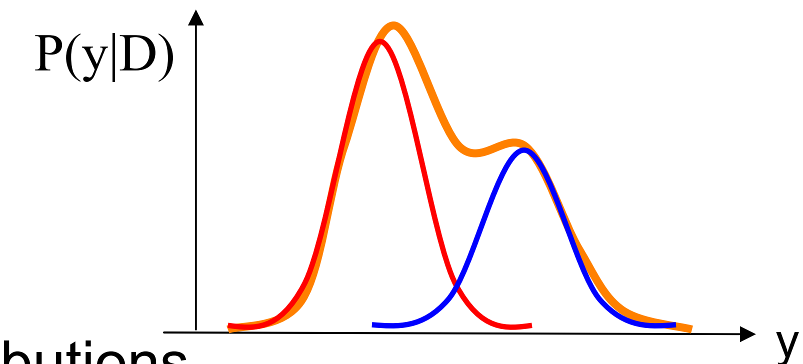
$$P(D|k)=\int L(D|m_k,k)P(m_k|k)\,dm_k$$

  - Bayes factor $B_{ij}= P(D|i)/ P(D|j)$
  - Model probability of model k:

$$P(k \mid D) = \frac{P(D \mid k)P(k)}{\sum_{j=1}^{N} P(D \mid j)P(j)}$$

Predictive distributions

- Bayesian model averaging: prediction for new y
  - $P(y|D) = \sum_k P(y|M_k,D)\, P(D|k)$



- Mixture distributions
  - Models/rock-types/properties
  - Most helpful way to manage heavy tails
- Often:
  - uncertainties *within* a model < uncertainty between models

Noise estimation

- Noise $\varepsilon \equiv (D\text{-}f(m))$

$\varepsilon = \varepsilon_{meas} + \varepsilon_{model}$

  - Often $\varepsilon_{model} > \varepsilon_{meas}$: we're prepared to live with approximate models
  - ε_{meas} usually stationary. Maybe biased. Roughly known.
  - ε_{model} often nonstationary. Correlated. Often biased. Usually unknown.
  - We blaze away: $\{\varepsilon, \sigma^2\} \sim N(0, \sigma^2)P(\sigma^2)$

- Inference of $p(\varepsilon)$ chief issue

  – $p(\varepsilon)$ strongly coupled to model dimensionality

  – Demands sensible priors in Bayesian formulation

    - *Simplest* form adds pieces like
      - $\mathrm{Log}(\Pi) \sim |D\text{-}f(m)|^2/(2\sigma^{2)}+n_D \log(\sigma^2)/2$

  - Absolute statements of model significance no longer possible

Marginal model likelihood estimators

- Full linearisation+conjugate priors, analytical
- Quadratures. Too hard as d=*dim*(m) gets large
- MCMC "Harmonic mean" averages:

$$P(D|k) = \int L(D|m_k,k)P(m_k,k)dm_k \qquad (1)$$

$$\hat{P}(D|k) = \left\langle \frac{1}{L(D|m_k,k)} \right\rangle^{-1}_{\Pi(m_k|D)(MCMC)}$$

Unstable as Nsamples→∞

- Laplace-Metropolis approx.

$$P(D|k) \approx \int e^{-f(m)}dm \approx (2\pi)^{d/2} \det\{\left.\frac{\partial^2 f}{\partial m_i \partial m_j}\right|_{\hat{m}}\}^{-1} e^{-f(\hat{m})} = (2\pi)^{d/2} |H|^{1/2} L(D|\hat{m})P(\hat{m})$$

- BIC asymptotics (eqn (1) does the "overfit" penalty automatically)

$$-\log(P(D|k)) \sim -\log(L(D|\hat{m})P(\hat{m}))) + \frac{d}{2}\log(n_D) + O(n^{-1/2})$$

Issues in prior formulation

- Lindley paradox (1957)
  - Models M_1 and M_2, with $d_2 > d_1$. Suppose they share (nested) parameters m with prior

$m_1 \sim N(0,\sigma^2 C_1)$

$m_2 \sim N(\{0,0\},\sigma^2 \,\mathrm{diag}\{C_1,C_2\})$

$P(\sigma) \sim 1/\sigma$ (Jeffrey's prior)

Then the Bayes factor is (Smith & Spiegelhalter 1980)

$B_{12} \sim |C_2|^{1/2} |X_2^T X_2|^{1/2} (1+((d_2-d_1)/(n-d_1)F)^{n/2}$

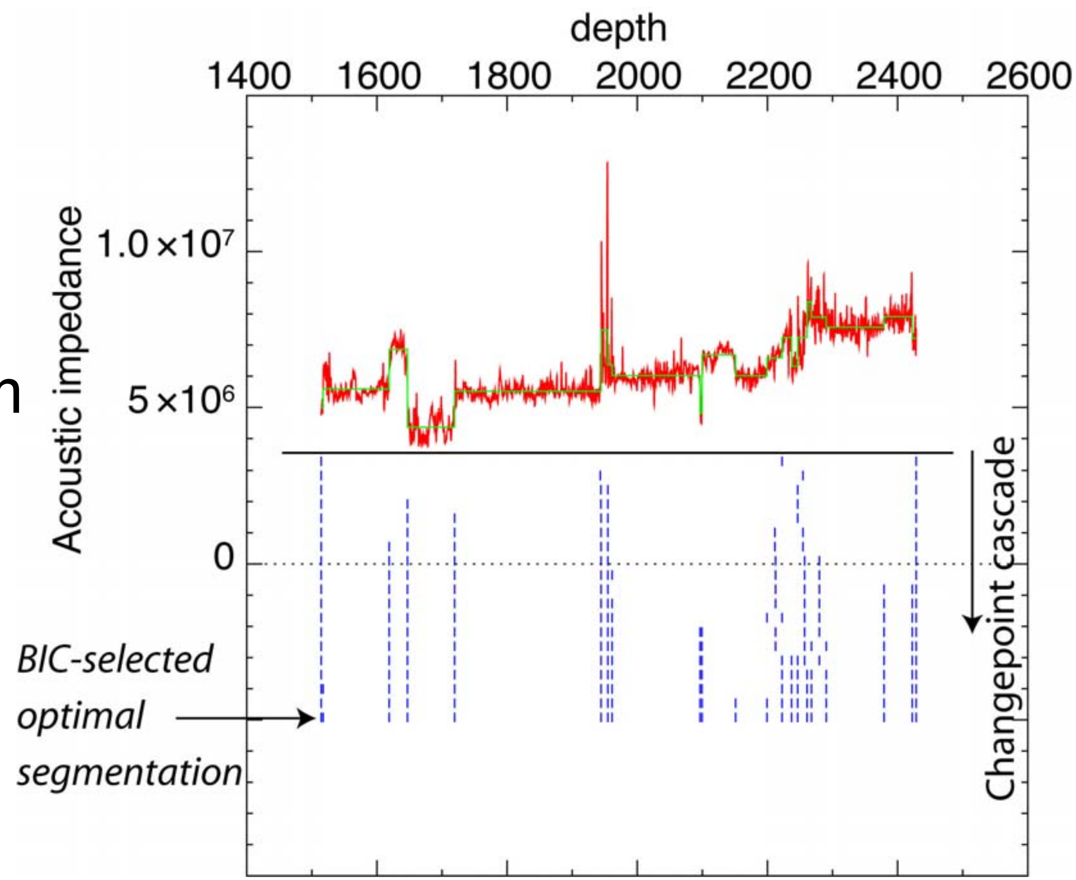
Always favours simplest model (1) as $|C_2| \to \infty$

(even for proper prior on m_2!)

- Setting of reasonable prior variances *very* important in comparing different dimensional models

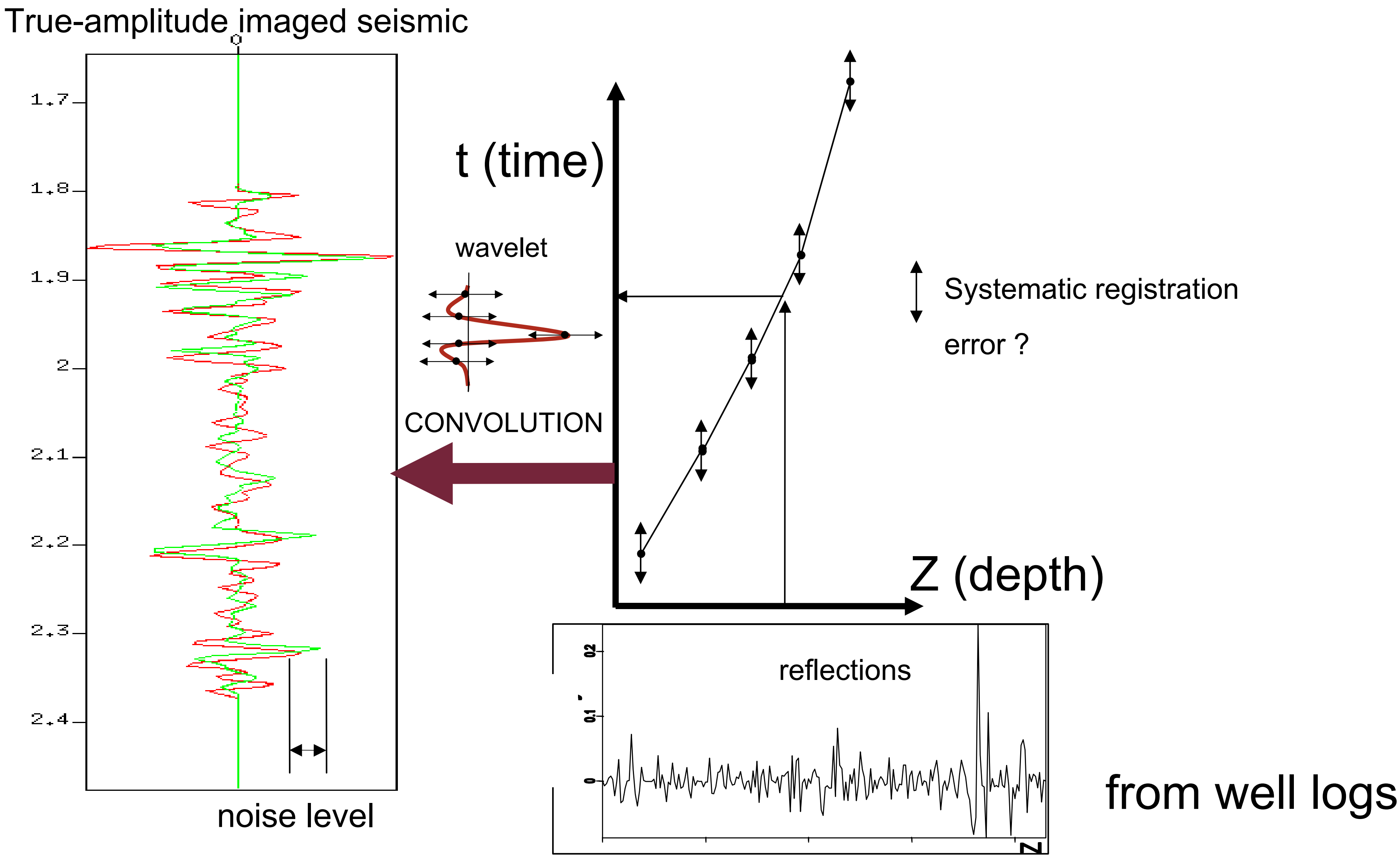
Segmentation/Blocking algorithms

- Changepoints a classic model-choice problem
- Maximum likelihood configurations for *k* changepoints τ_j accessible from dynamic programming $O(kn^2)$ (D.M.Hawkins, Comp. Stat. & Data Analysis 37(3), 2001)
- Uses:
  - Thinning reflectivity sequence
  - Optimal checkshot choices



$$-\log(L(y,m,\tau)) = \frac{1}{2}\sum_{j=1}^{k}\sum_{i=\tau_{j-1}+1}^{\tau_j}(y_i - m_j)^2/\sigma^2 + \frac{n}{2}\log(\sigma^2) + BIC$$

Wavelet extraction problem



Real wavelets and effective wavelets: Ricker 1953 (dynamite)

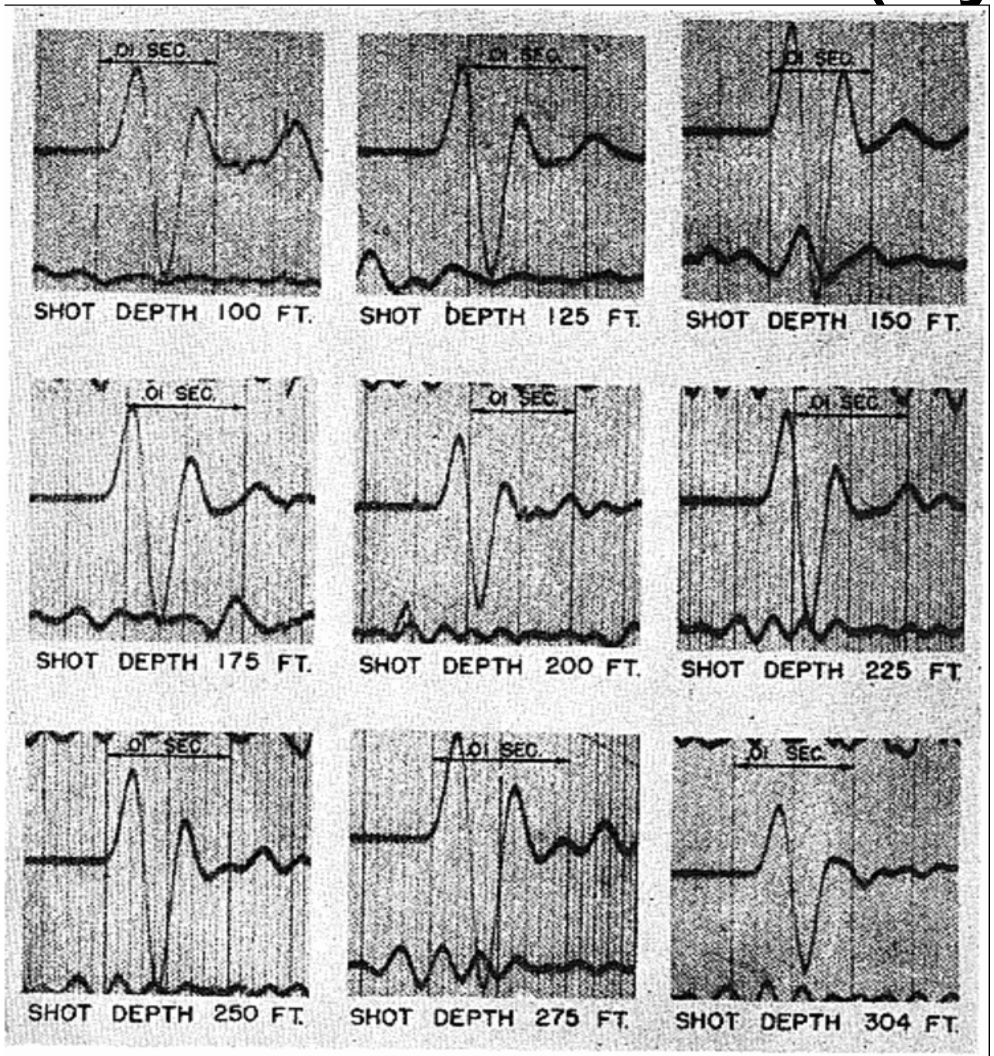


FIG. 5. The primary seismic disturbance, velocity type, as observed in Pierre shale.

Motivations

- Wavelet extraction a critical process in inversion studies
- Commercial codes don't (or wont) tell you
  - Noise estimates
  - Non-normal-incidence wavelets
  - Time to depth or positioning errors
  - *May* struggle with
    - Multiple wells
    - Deviated wells
  - wavelet uncertainties

Wishlist

- Wavelet coefficient extraction (plus errors)
- Wavelet length estimation (plus errors)
- Noise estimation (plus errors)
- Petrophysics model choices
- Time to depth adjustments(plus errors)
  - Integrates checkshots and sonic logs
- Positioning adjustments (plus errors)
- Multi-well
- Multi-stack

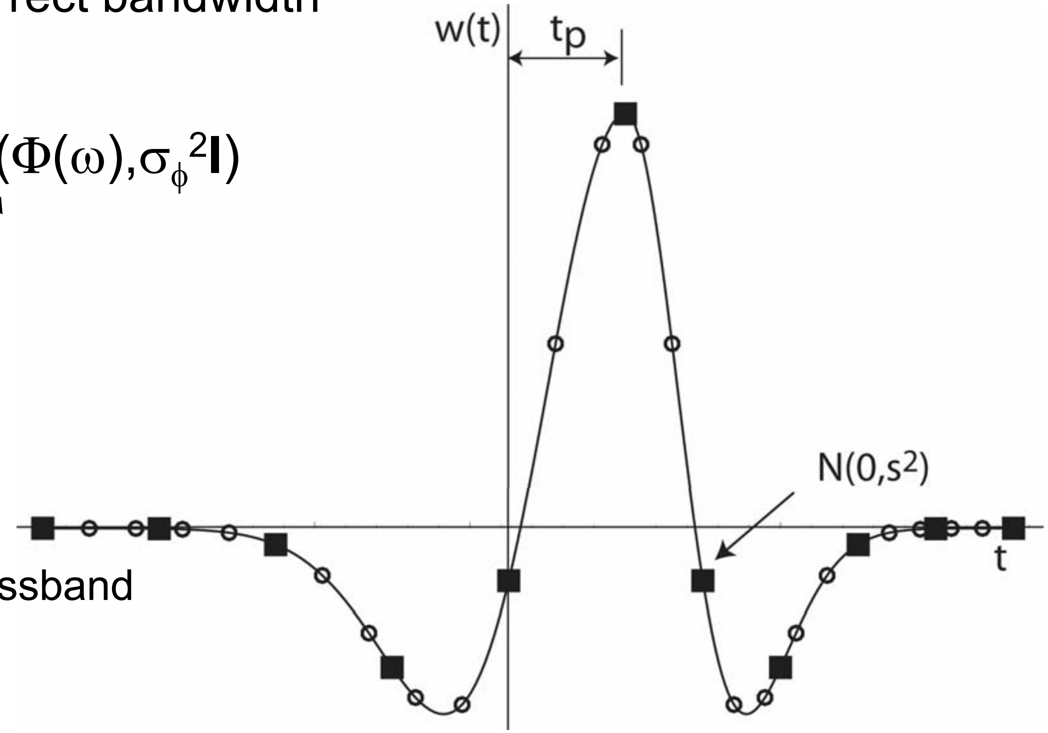
Wavelet Parameterization (1)

- Wavelet coefficients m_w
  - Nyquist-rate knots on cubic splines with zero endpoints & derivatives
  - Ensures decent tapering & correct bandwidth
  - Span ? (see later)

- Prior $P(m_w)=N(\mathbf{0},s^2\mathbf{I})N(t_p,\sigma_p^2)N(\Phi(\omega),\sigma_\phi^2\mathbf{I})$

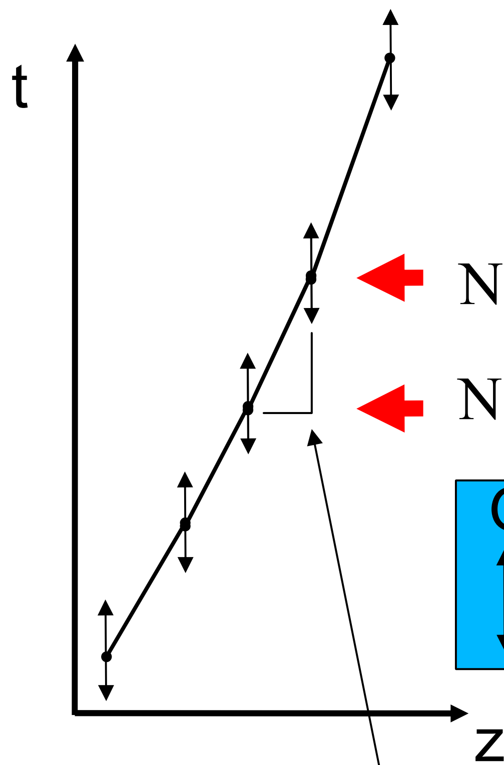
Optional timing prior

Optional phase prior in passband



Checkshot Parameterization (2)

- Time to depth parameters



$N(\mathbf{t}_{m,i},\sigma_{m,i}^{2})$ Markers (geological identifications)

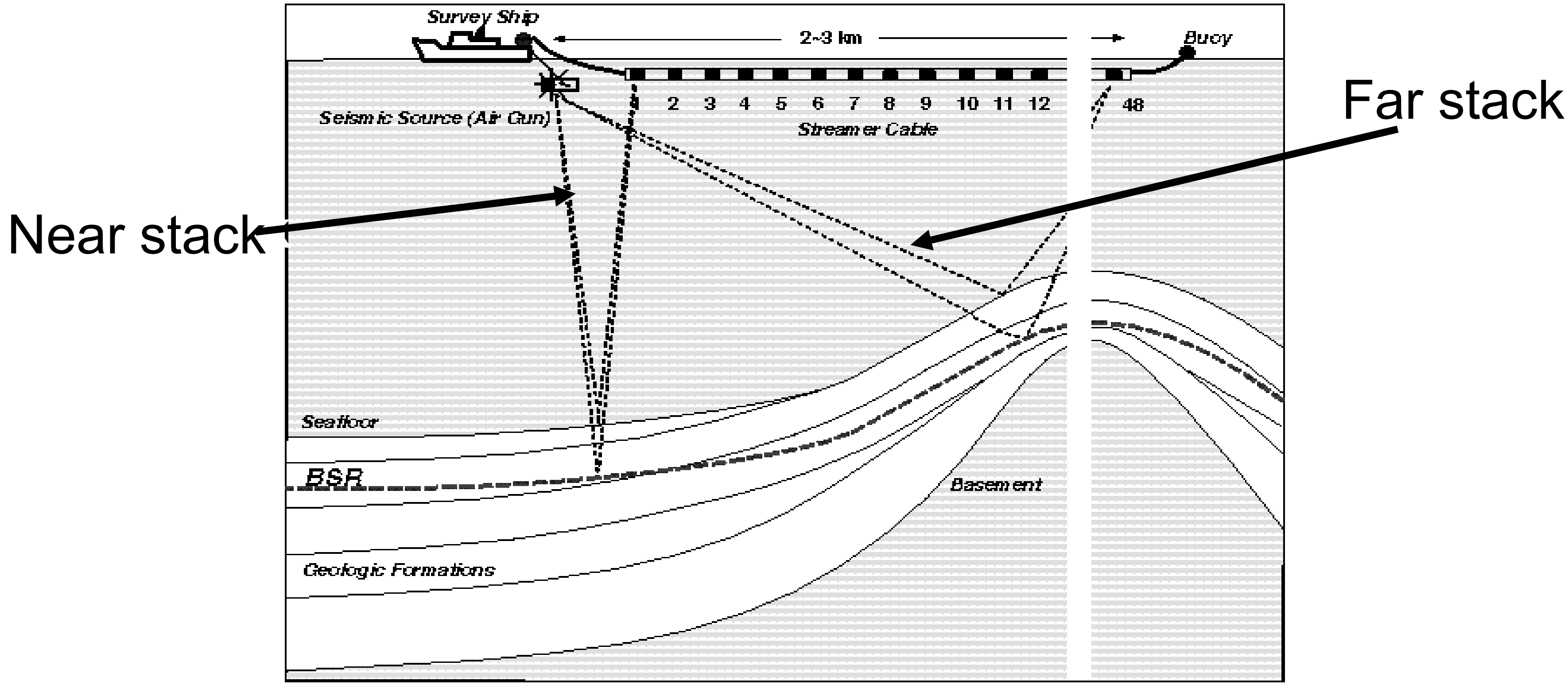
$N(\mathbf{t}_{c,i},\sigma_{c,i}^{2})$ Checkshots (e.g. from VSP tools)

Global (well specific) time shift

$t_g \pm \sigma_{tg}$

$v_{int}(z_{i+1}-z_i\,,t_{i+1}-t_i) \sim N(v_{log\text{-}average},\ \sigma_v^{2})$

Petrophysics parameterizations: Multi stack



Petrophysics Parameterizations(4): AVO effects (Linearized Zoeppritz)

- General approximate p-p reflection coefficient

normal incidence   angle unknown?

$$R=\boxed{\tfrac{1}{2}(\Delta\rho/\rho + \Delta v_p/v_p)} + B(\boxed{\theta^2(\tfrac{1}{2}\Delta v_p/v_p - 2v_s^2(\Delta\rho/\rho+2\Delta v_s/v_s)/v_p^2)}$$

shear dependent

$$+\boxed{\theta^2)\Delta\delta/2}$$

Thomsen anisotropy parameters:

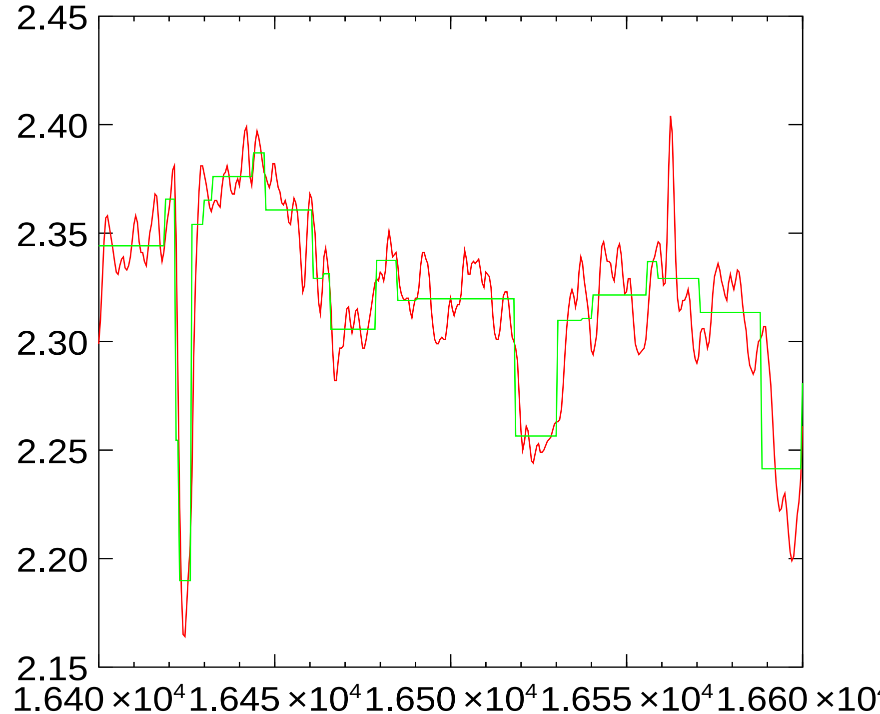
- Stack angle $\theta^2 = v_p^2/(V_{stack}{}^4 T_{stack}{}^2/\langle X_{stack}{}^2\rangle)$
- R computed at block boundaries: v_p, v_s, ρ are Backus-upscaled segment properties (i.e. error-free log data)
- Anisotropy: $\delta \sim N(\delta_f, \sigma_f^2)$, f=lumped facies parameter label (from classification of segment)

Anisotropy issues

- Processing issues for AVO
  - Many potential amplitude effects at wider offset
    - Absorption/scattering/spreading/acquisition footprints, etc.. etc.. etc
  - Ch. 4.3 “Quantitative seismic interpretation” Avseth, Mukerji, Mavko.

Preprocessing: Impedance Blocking

- Based on segmentation of p-wave impedance (ρv_p)
  - Max Lik. methods $O(N^2)$, or
  - Blended hierarchical stepwise segment/aggregate method ($O(N\log(k))$)
  - ref: D.M.Hawkins, Comp. Stat. & Data Analysis 37(3), 2001
- Increases speed



Optimisation of Bayesian Posterior

- $m = \{m_{wavelet}, m_{checkshot}, m_{registration}, m_{petrophysics}\}$
- Maximise $\prod \propto \exp(-\chi^2/2)$

$$\chi^2 = \sum_{t,stacks,wells} (w(m) * R_{eff} - S)^2 / \sigma_s^2$$ "good synthetic"

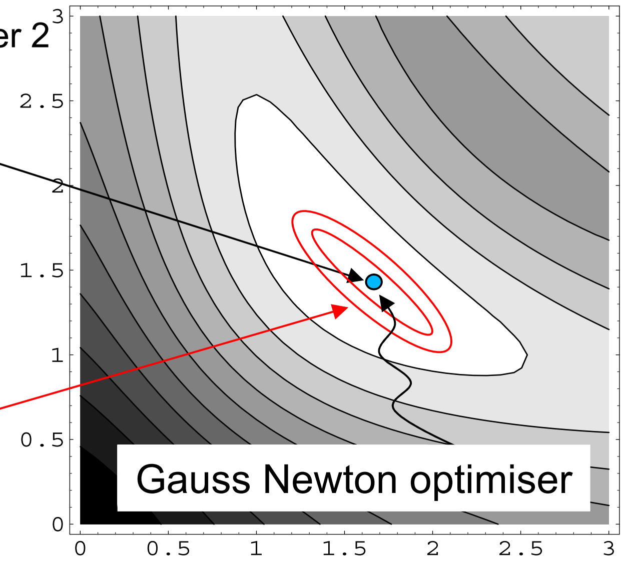
$$+ (d+1) \sum_{stacks} \log(\sigma_s)$$ "noise normalisation"

$$+ \sum_{intervals,wells} (v_{int}(m) - \langle v_{int} \rangle)^2 / \sigma_v^2$$ "prior: interval velocities c.f. sonic logs"

$$+ (m - \langle m \rangle)^T C_p^{-1} (m - \langle m \rangle)$$ "prior on checkshots, wavelet coeffs"

$$+ \text{wavelet timing/phase prior}$$ "e.g. constant phase"

Optimisation, Hessians, Approximate uncertainties



Maximum aposteriori estimates

Quadratic approximation to uncertainties
- (1) direct from F.D. Hessian or
- (2) from Bayes' update formula
- $C=(Cp^{-1}+X^{T}C_{D}^{-1}X)^{-1}$

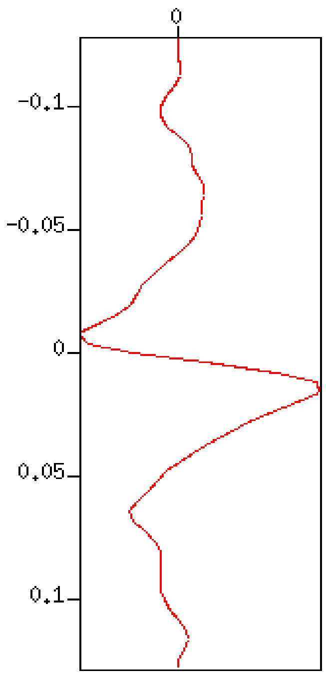
log(posterior) surface

- Gauss-Newton, BFGS methods: see e.g. Nocedal & Wright "Numerical Optimization"
- $O(d^2)$ x O(forward model cost) x N(models)
- Typical dimensionalities: d=10-100, n = 100-1000

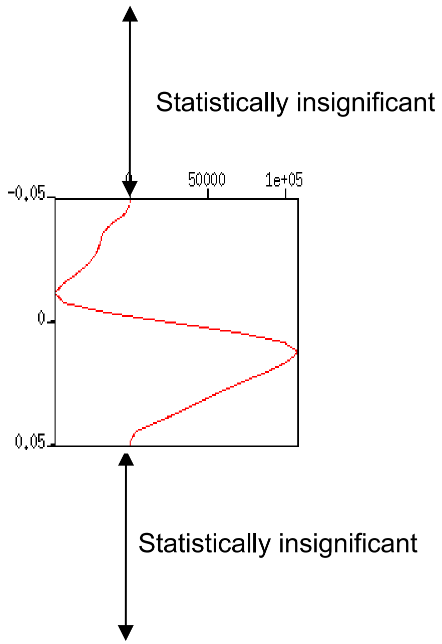
Optimisation outputs

- Maximum aposteriori parameters
  - SU wavelets, ascii time-to-depth files, visualization dumps etc.
- Covariance diagonals (parameter uncertainties)
- Posterior model probabilities (wavelet span, checkshot choice etc)
- stochastic wavelets from the posterior

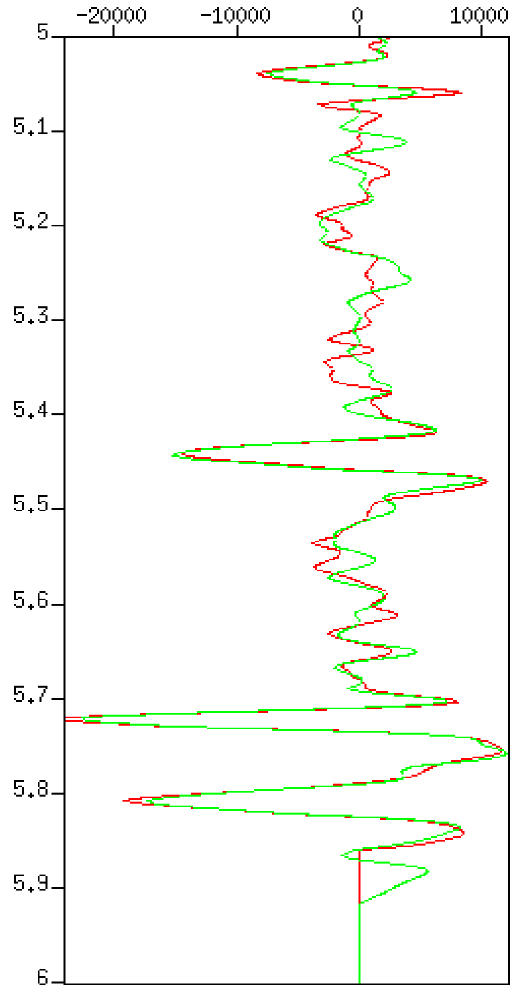
Generic aspects of well-ties

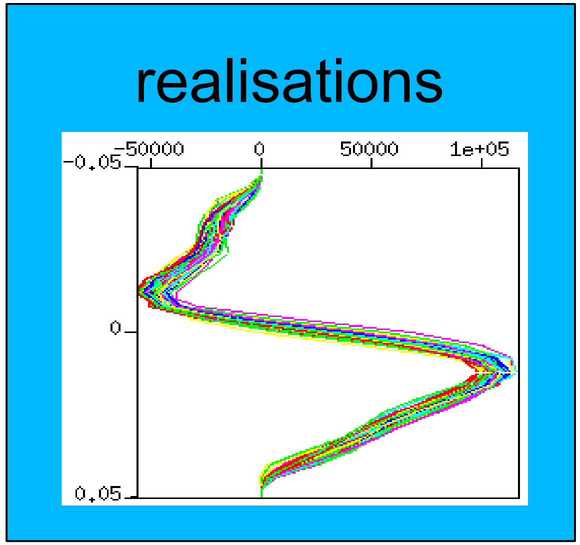


Typical commercial extraction

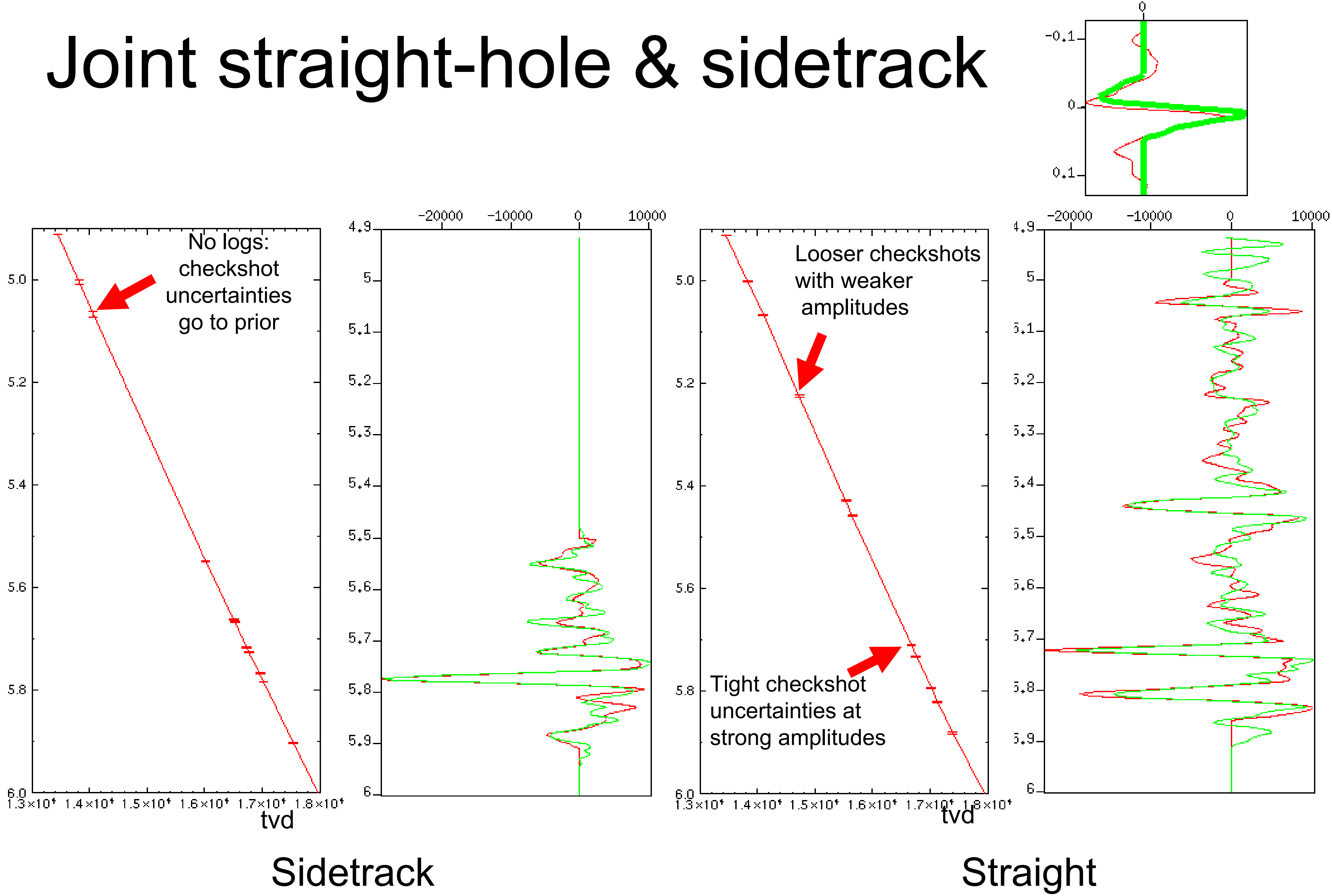


Bayes ML wavelet

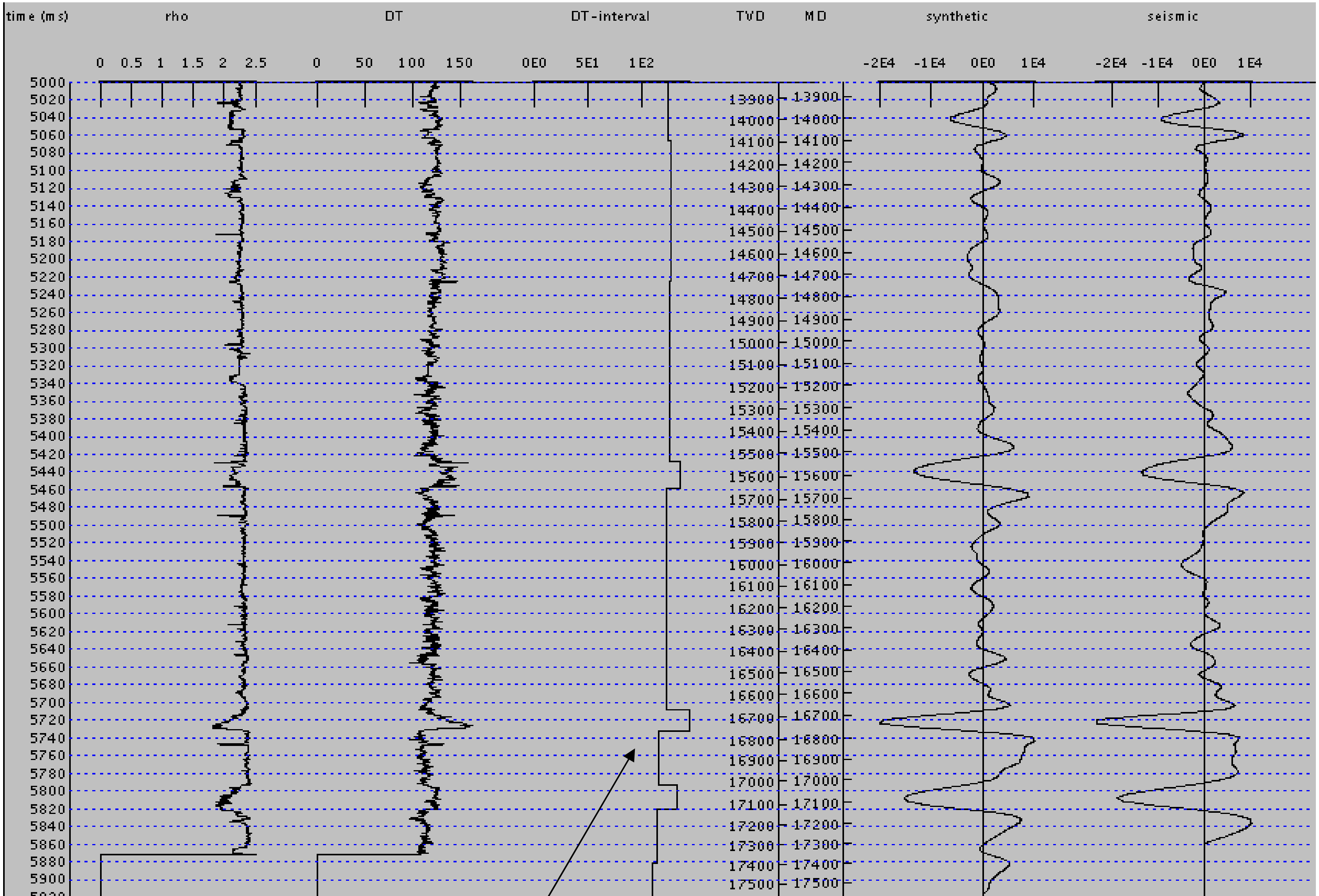




Joint straight-hole & sidetrack

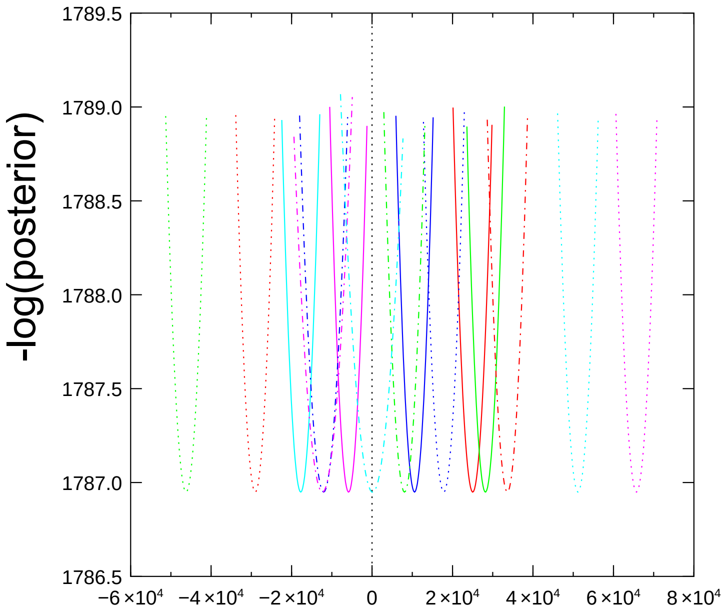


Straight hole detail with logs

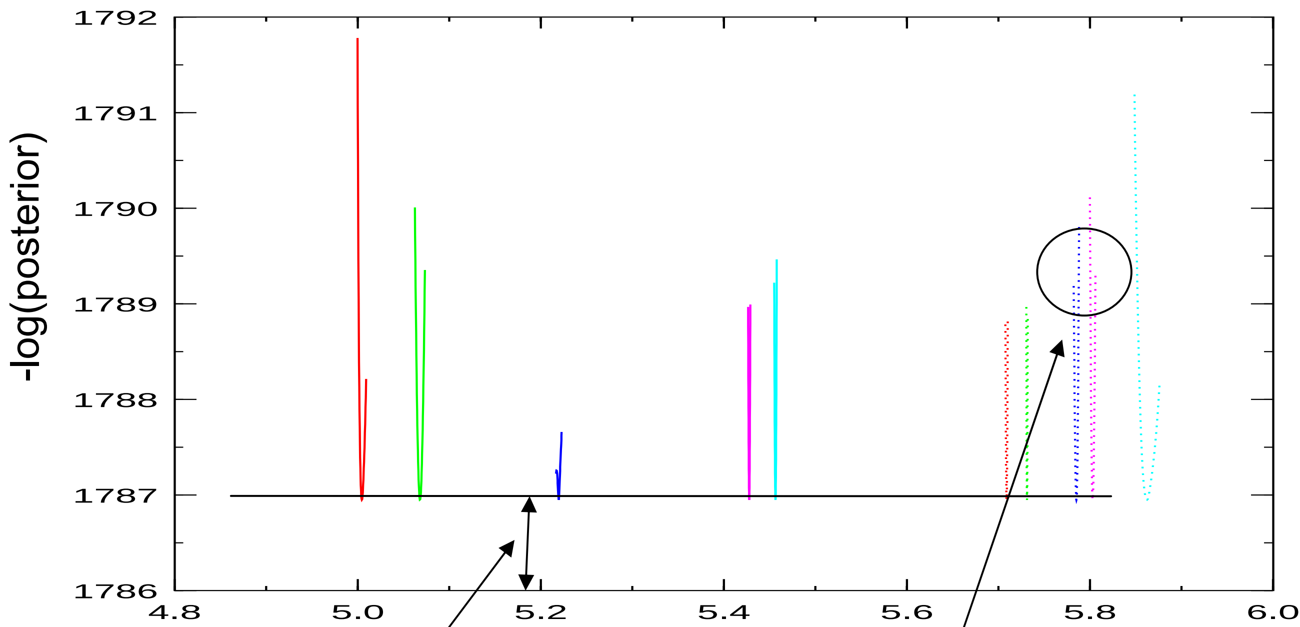


Interval sonic velocities constrained to within 5% of log average

Parameter uncertainty cross sections



wavelet coefficients



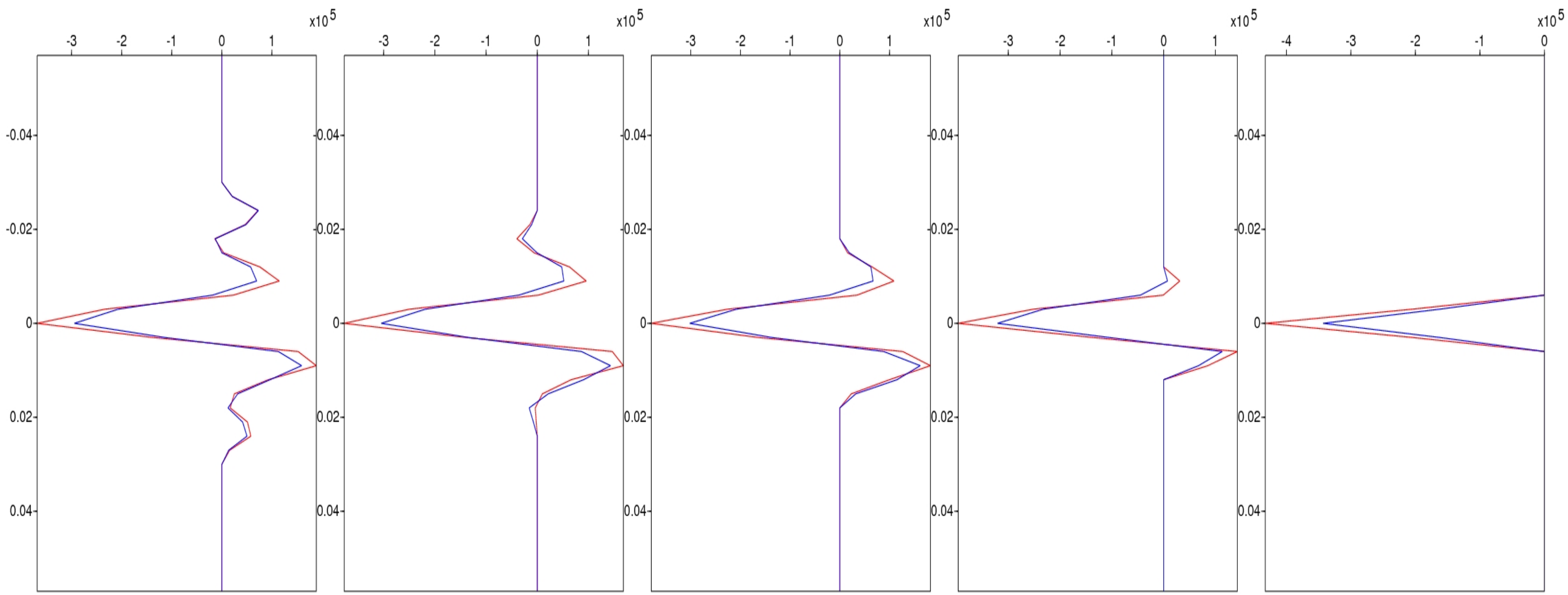
checkshot knots

One standard deviation

repulsion to keep interval velocities sane

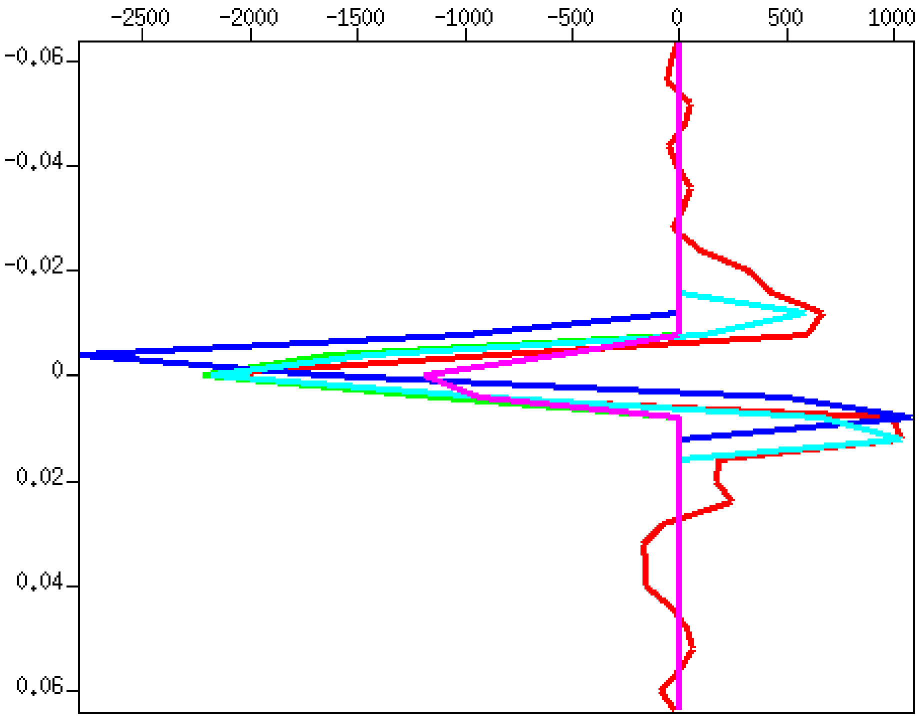
General effect of allowing more freedom

- Wavelets sharpen, noise reduces

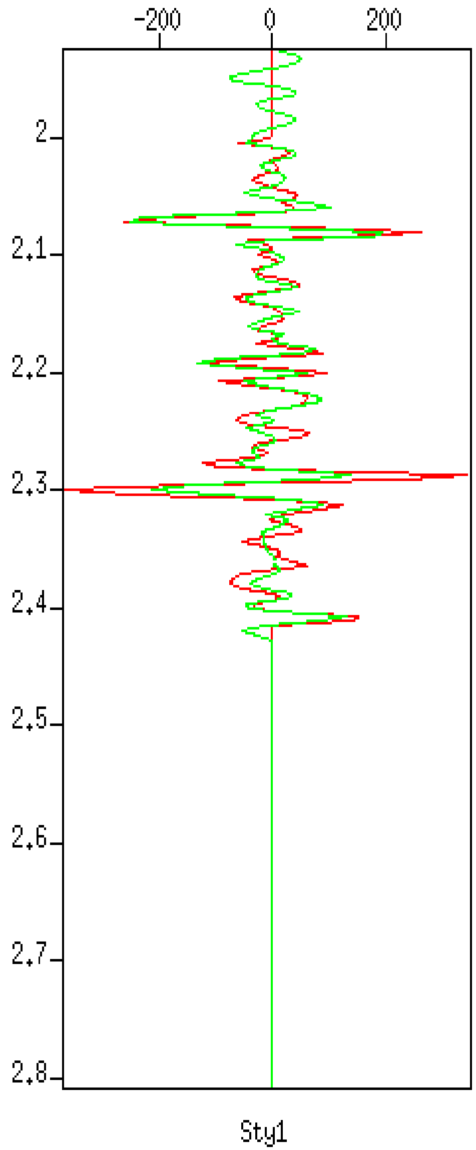


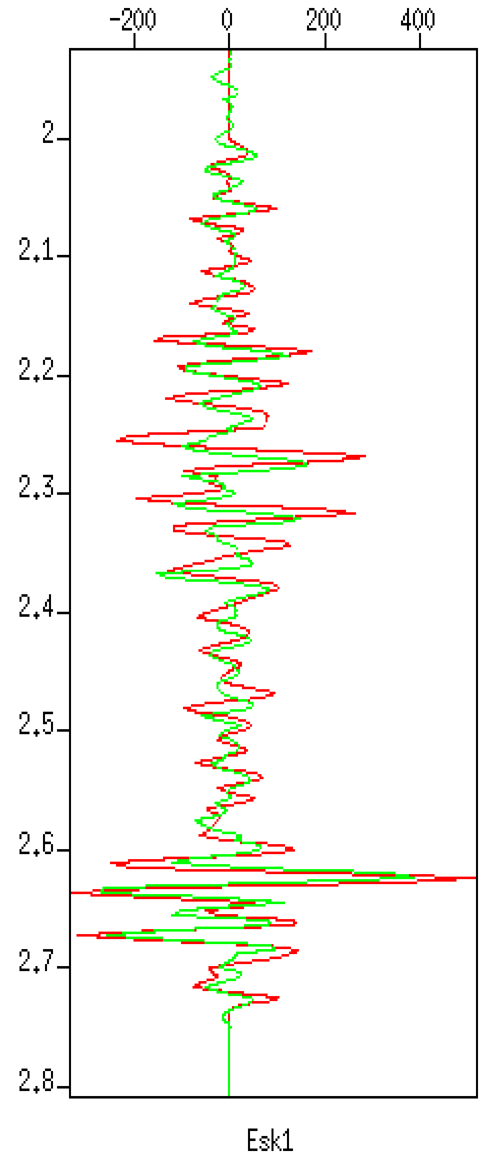
Cross well issues: Wavelets and Ties

Wavelets from various extractions



Synthetics/traces from joint extraction





Relations to Cross-validation

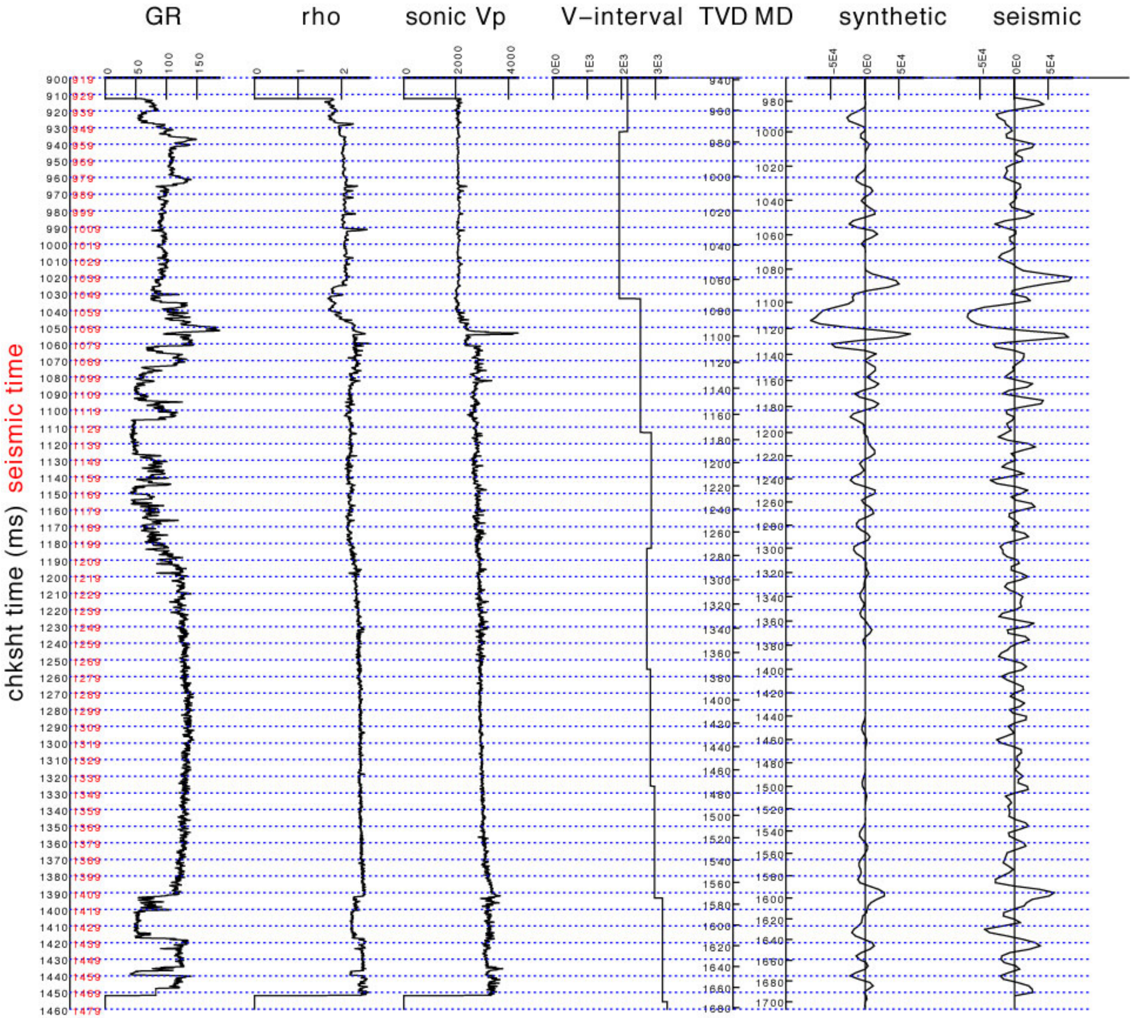
- Leave-one-out pseudo Bayes factor

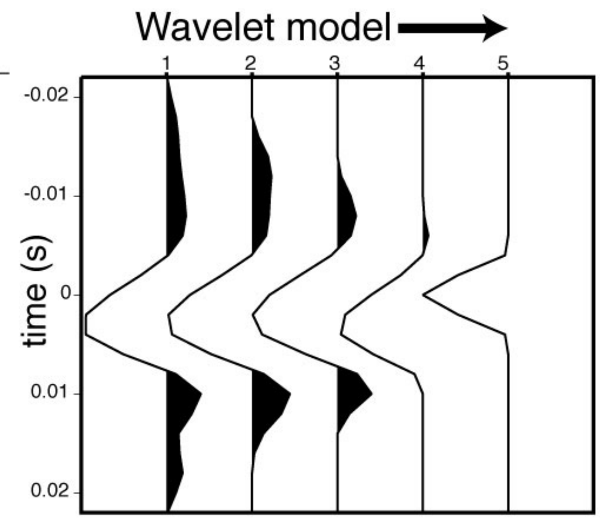
$$B_{12(cross-val)} = \prod_{i=1}^{n_D} \frac{p(d_i \mid d_{\{j \neq i\}}, M_1)}{p(d_i \mid d_{\{j \neq i\}}, M_2)}$$

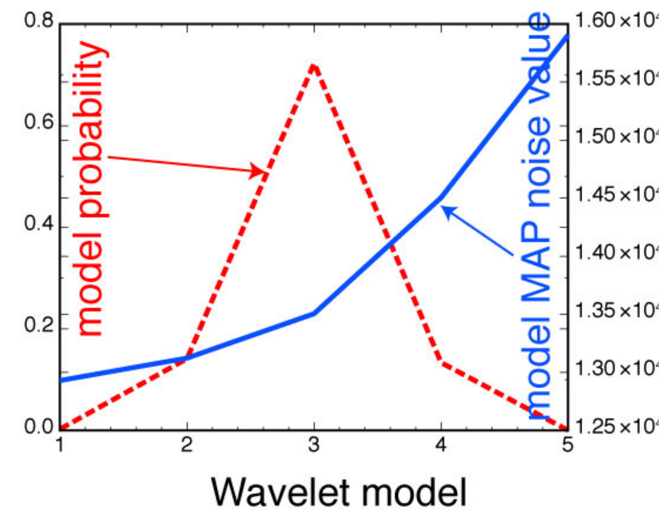
- Has asymptotics of Akaike information criterion
  - Log(P(D|k)) += d (c.f. Bayes info. crit. = (d/2)log(n_D)
  - (less severe on richer models)

3 Model Selection examples

(1) Simple Wavelet-length choice

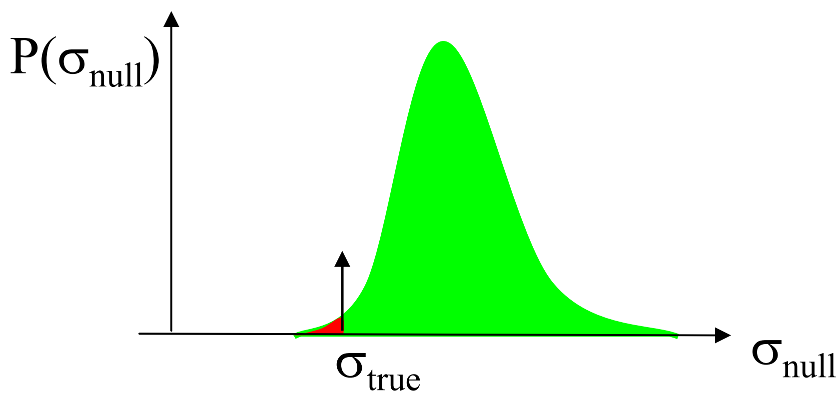




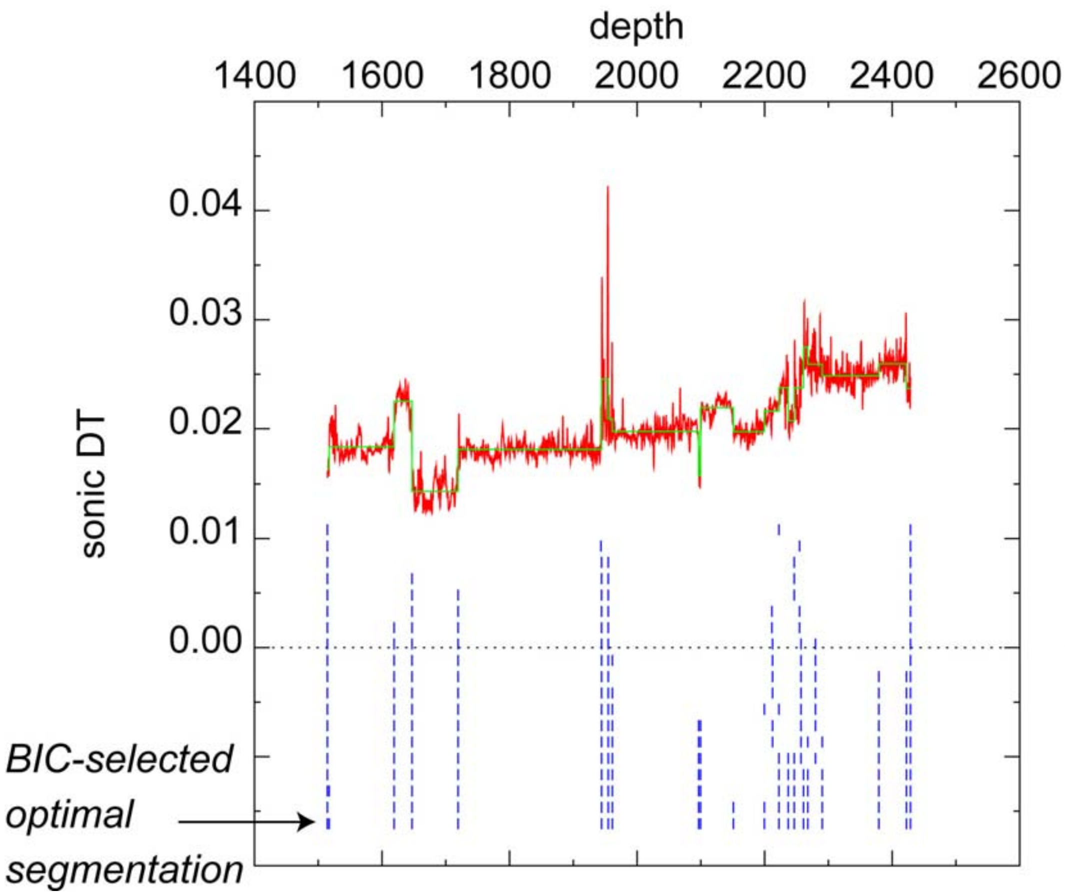


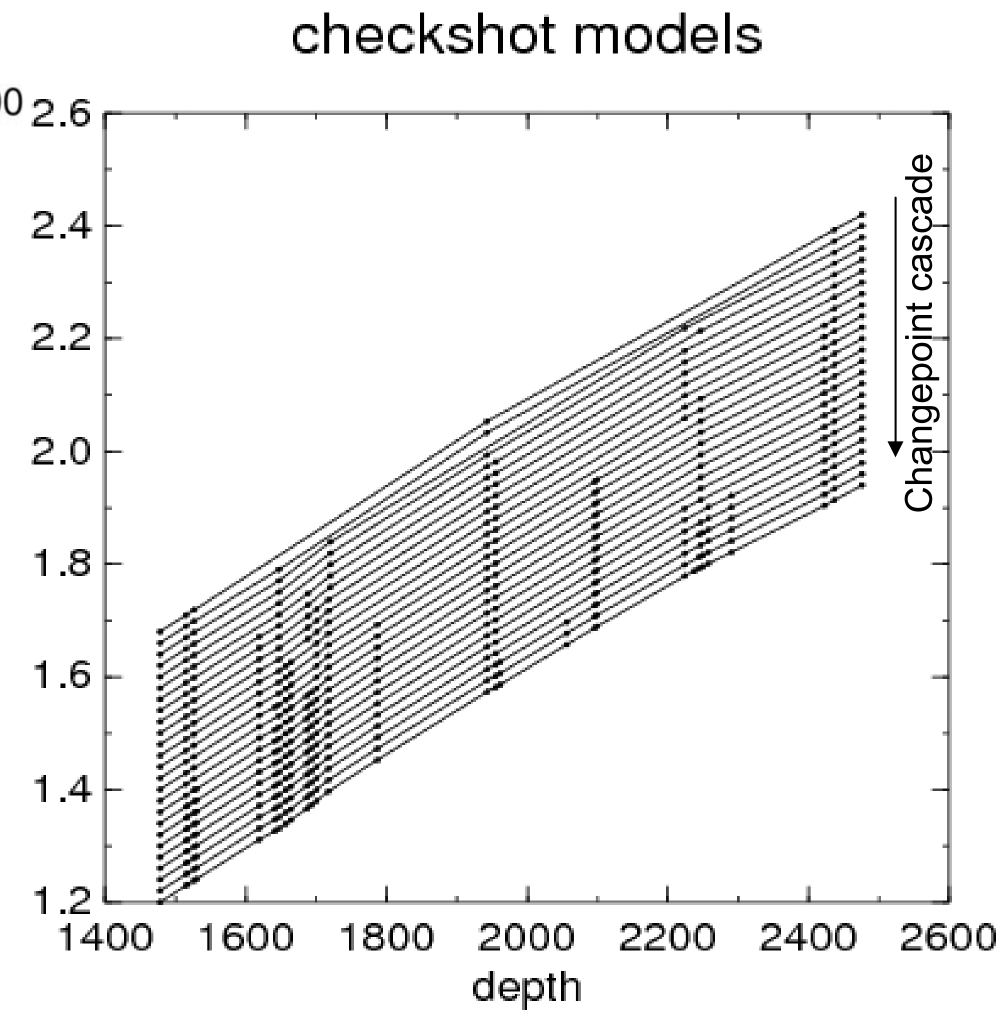
Absolute significance tests

- MAP Wavelet length *not* shortest model
- MAP noise level exceptional on null-hypothesis test
  - Perform ensemble of extractions on fake seismic data with matched bandwidth and univariate statistics:

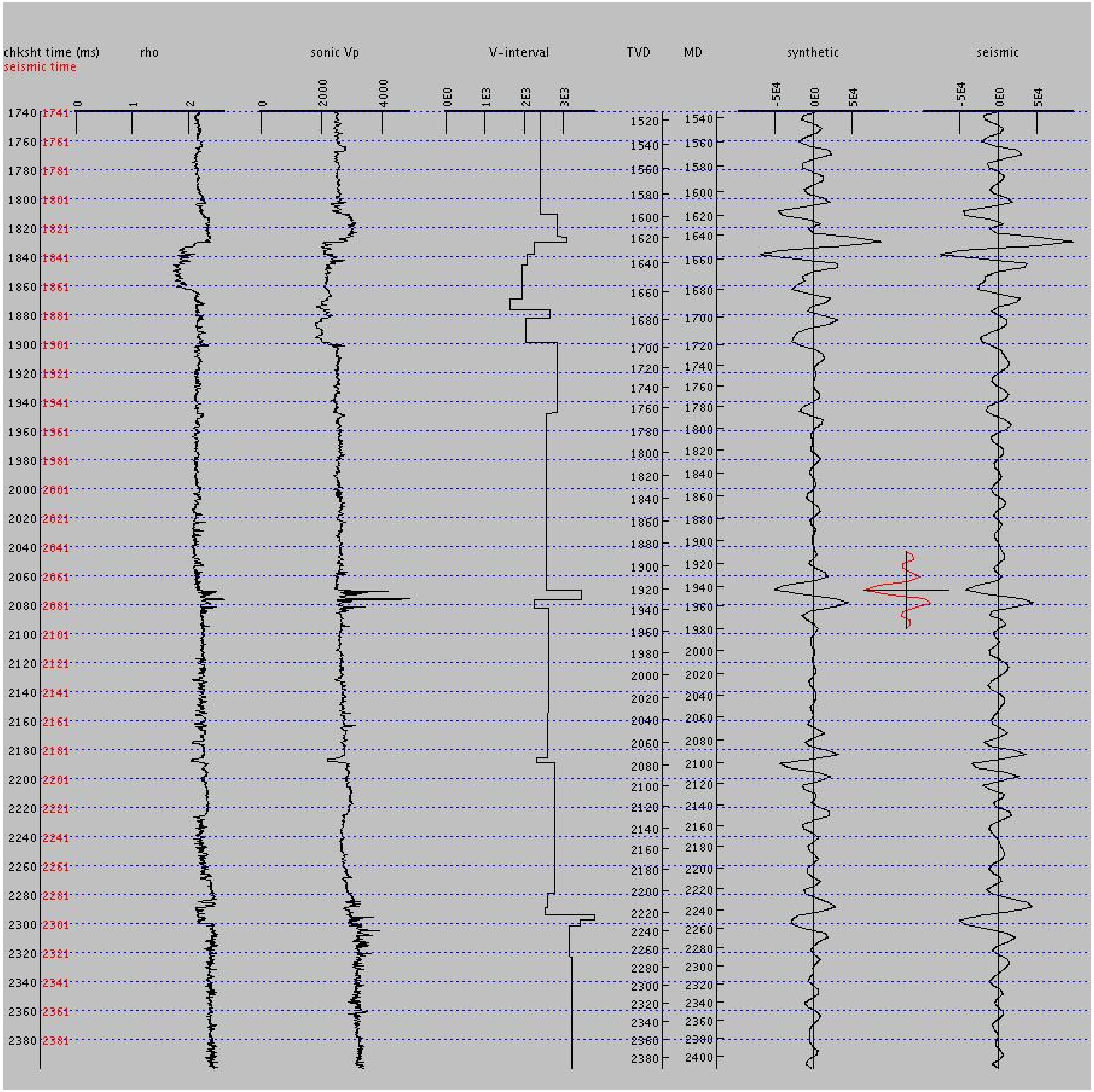


(2) Checkshot choice (25 models)

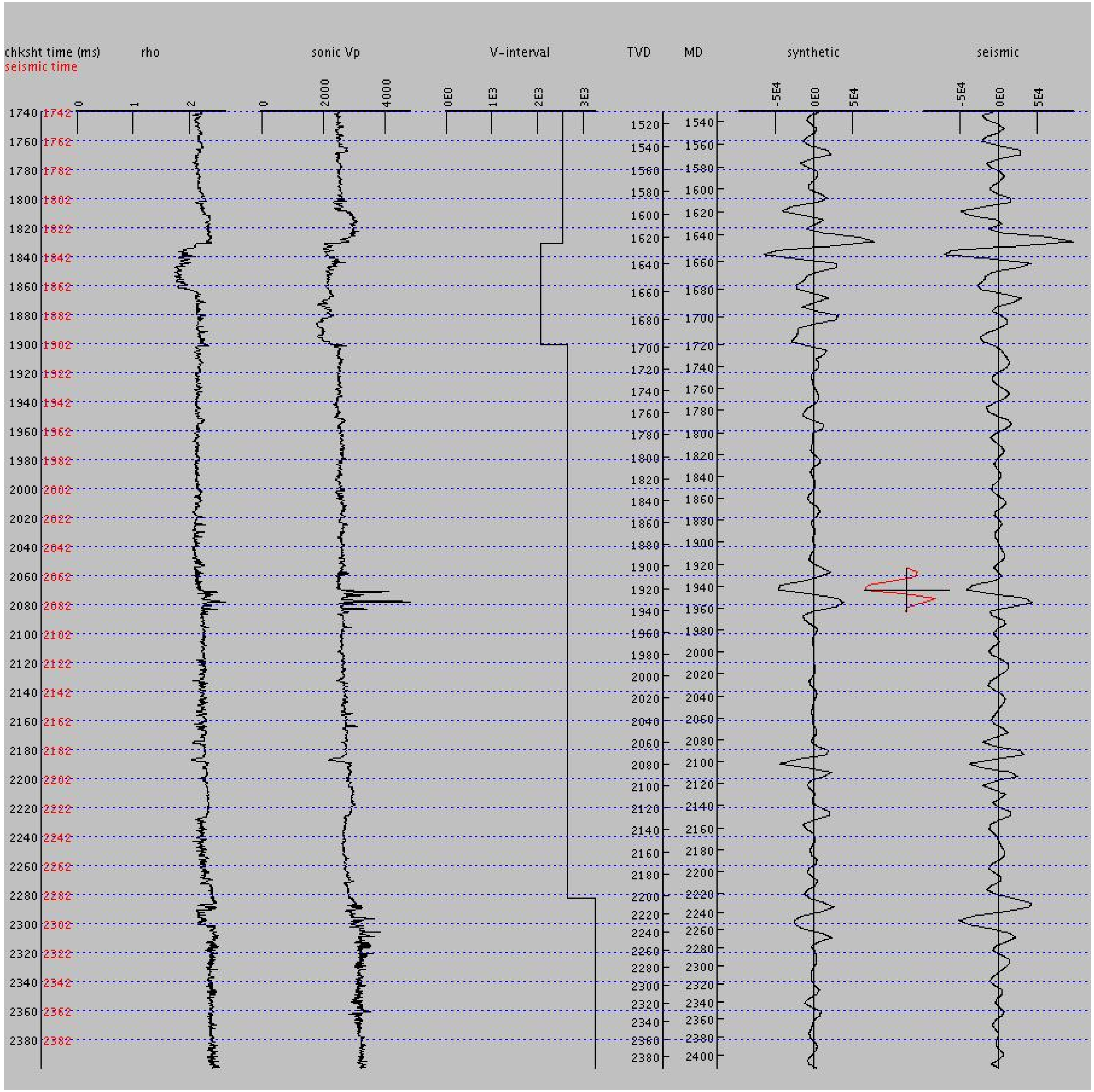




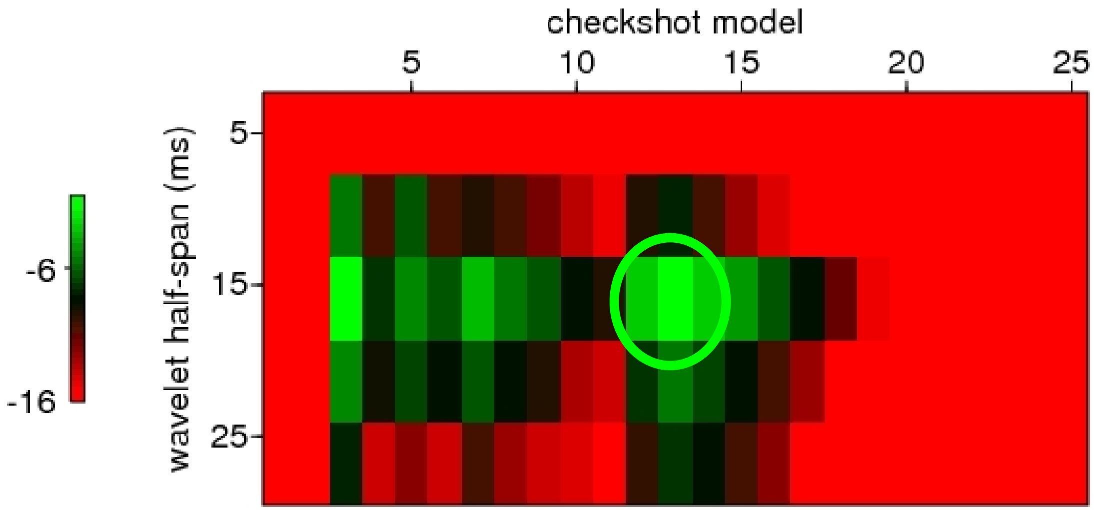
Richest model



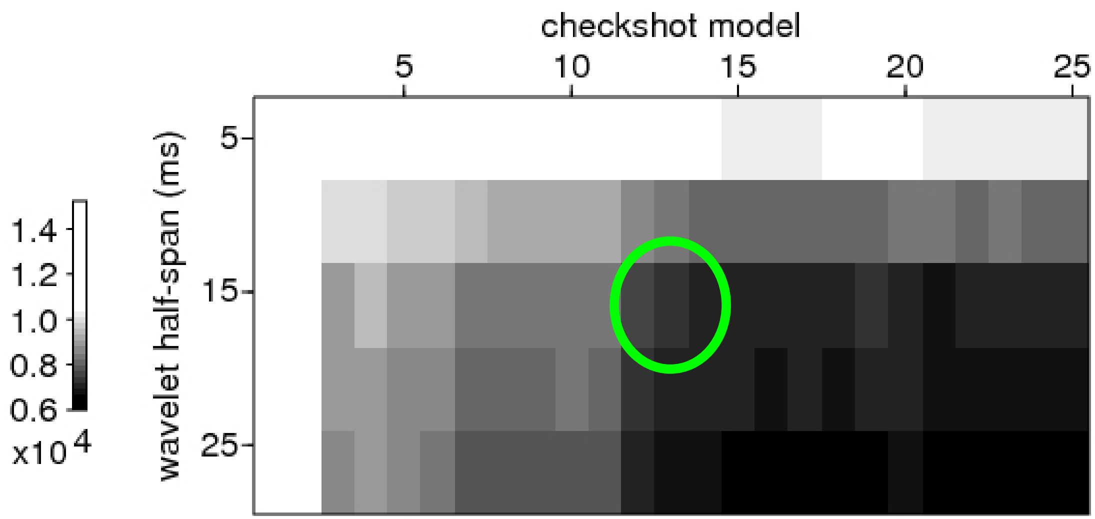
Max. aposteriori model



Model probabilities and noise



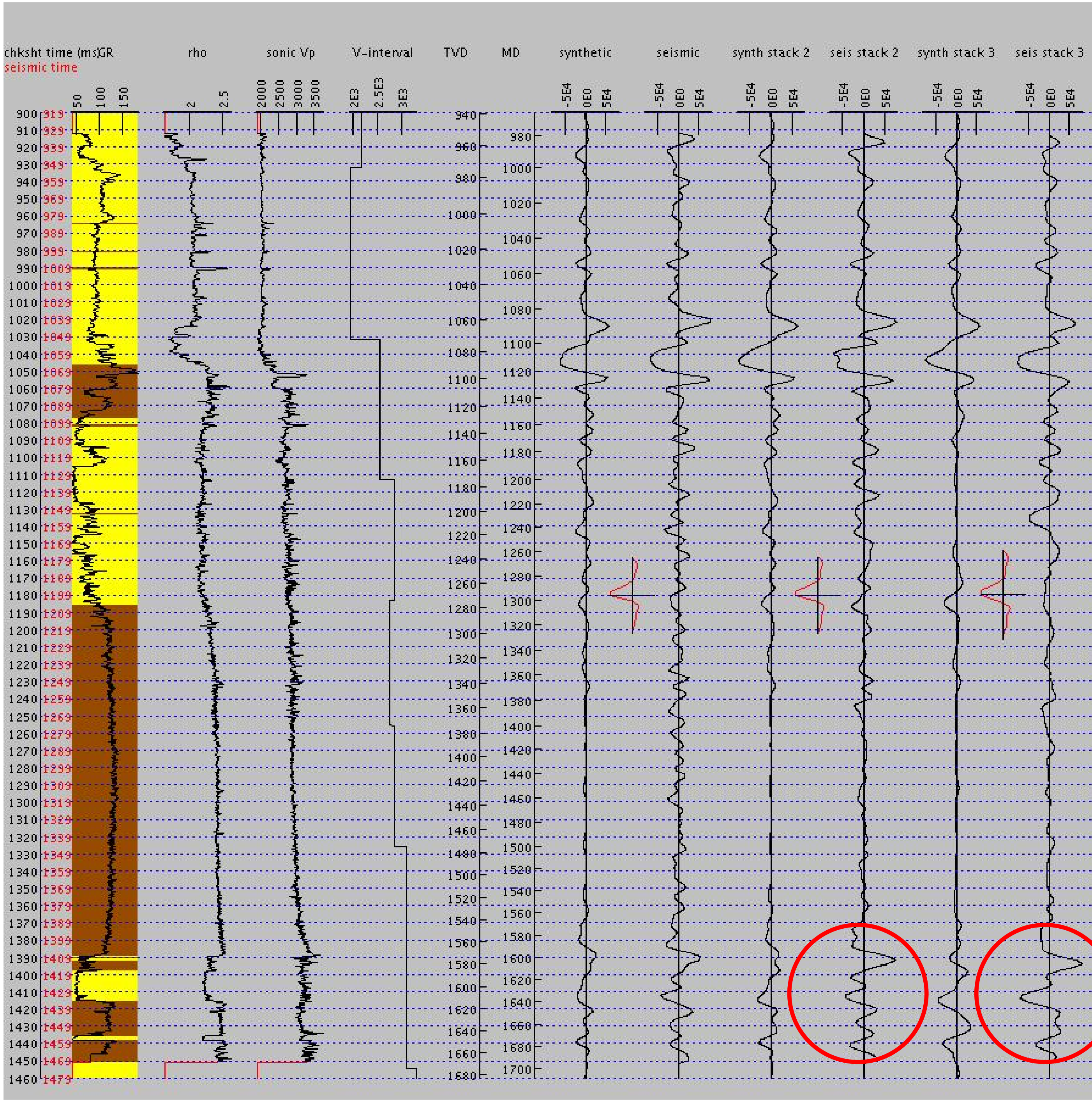
log(model MAP probability)



noise rms

(3) Naïve facies-driven anisotropy choice

$\delta \sim 0.15$



Conclusions

- Model selection is desirable in nearly all inverse problems
- Marginal model likelihoods can be tricky
  - Careful construction of prior
  - Efficient (differentiable) forward models
  - Efficient optimisers
- Bayesian model selection is quite “opinionated”. Very Occamist: severe on overparameterised models
- Selection is a doable problem for wavelet extraction. Produces
  - Often compact wavelets
  - Simplified checkshots
  - Significant potential for missing-data (petrophysics) problems

O me, the word 'choose!' I may neither choose whom I would nor refuse whom I dislike.

Portia, The Merchant of Venice.

More info:

- www.google.com/search?q=James+Gunning+CSIRO
  - “Wavelet extractor” open source code (java), demos, papers.
  - Part of “Delivery” suite (Bayesian seismic inversion)
- Computers and Geosciences **32** (2006)

Data-positioning uncertainty (1)

- $y(m) = f(m)$ an unusual regression problem
- Simple toy problem; fitting a line $y=a+bx$:
  - $y \sim X.m + e$,   Jeffrey's prior on $e \sim N(0,\sigma)$
  - Priors:
    - $P(\sigma) \sim 1/\sigma$
    - $P(m|\sigma) \sim N(0, g\,\sigma^2(X^TX)^{-1})$ ("Zellner")
  - Posterior marginal

    $$\prod = \int L(y|m,\sigma)P(m|\sigma)P(\sigma)dmd\sigma \sim (1+g(1-R^2))^{-(n-1)/2}$$
  - Where R^2 is usual coeff. of determination
- For two samples of “random” y
  - $B_{12} \sim ((1+g(1-R_1^2))/(1+g(1-R_2^2)))^{-(n-1)/2}$

Data-positioning uncertainty (2)

- Fluctuation in B_{12} for problem with
- y=1+x+(25% fixed noise) + (5% fluctuations)