Improving the spatial resolution of effective elastic thickness estimation with the fan wavelet transform

J. F. Kirby*1 and C. J. Swain2

* Corresponding author.

1 Department of Spatial Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia. Tel: +61 8 9266 7701; fax: +61 8 9266 2703; email: j.kirby@curtin.edu.au.

2 Department of Spatial Sciences, Curtin University, GPO Box U1987, Perth, WA 6845, Australia. Email: c_swain@wt.com.au.

Abstract

We show here a simple technique to improve the spatial resolution of the fan wavelet method for effective elastic thickness ($T_e$) estimation that we have previously developed. The technique involves reducing the number of significant oscillations within the Gaussian window of the Morlet wavelet from approximately five to three or fewer (while making an additional correction for its no-longer-zero mean value). Testing with synthetic models and data over South America indicates that the accompanying reduction in wavenumber resolution does not seriously affect the accuracy of the $T_e$ estimates. Comparison against the more widely-used multitaper Fourier transform approach shows that the enhanced wavelet method not only improves upon the multitaper method’s spatial resolution, but also is computationally much faster and requires the arbitrary variation of only one parameter compared to three for the multitaper method. Finally, we present a modified method to compute the predicted coherence using the multitaper method that, while not improving its spatial resolution, does improve the bias of recovered $T_e$ estimates.
1. Introduction

Since the mid 1990s, many spectral methods have been developed that are able to reveal spatial variations in the effective elastic thickness ($T_e$) of the lithosphere, whereas before only single estimates for a region were possible (e.g., Zuber et al., 1989). For example, Lowry and Smith (1994, 1995) applied the maximum entropy method (MEM) to the US Basin and Range province, Simons and van der Hilst (2002) used a multitaper Fourier transform in Australia (although they mapped the variations of coherence transition wavelength – see below), Braitenberg et al. (2002) used a space-domain convolution approach in the Alps, and Stark et al. (2003) developed a wavelet transform technique (based on Gaussian tensor wavelets) and applied it to southern Africa. Subsequently, Kirby and Swain (2004) developed an alternative wavelet method, this time based on superposed Morlet wavelets and called the ‘fan’ wavelet transform; this method has been applied to, for example, Australia (Swain and Kirby, 2006), and the Canadian shield (Audet and Mareschal, 2007).

All of the above-cited methods use variations of the coherence method of Forsyth (1985), in which the coherence between observed Bouguer gravity anomalies and topography is computed in the wavenumber ($k$) domain through:

$$\gamma^2(|k|) = \frac{\langle GH\rangle \langle GH\rangle^*}{\langle GG\rangle \langle HH\rangle^*}$$  \hspace{1cm} (1)
where $G$ and $H$ are the tapered Fourier or wavelet transforms of Bouguer gravity anomaly and topography, respectively, the * indicates complex conjugation, and the angular brackets indicate an averaging process (see Kirby and Swain, 2009, for a discussion on averaging). This observed coherence is then compared against the predictions of a thin elastic plate model to find the best-fitting $T_e$. Theoretical Bouguer coherence curves have a characteristic shape, maintaining high values (close to 1) at long wavelengths, indicative of isostatic compensation, and low values (close to 0) at short, indicating mechanical support. At some “coherence transition” wavelength in between there is an often-sharp rollover from 1 to 0: the longer the wavelength of this rollover, the stronger the plate.

In the multitaper (MT) implementation, the study area is divided into overlapping windows of one, or a few, fixed window sizes, the coherence computed (via a multitaper Fourier transform) and inverted in each window. The resolution band is hence prescribed by, among other parameters, the choice of window size. The wavelet method, however, convolves a range of scaled wavelets with the whole data set to map and invert the coherence at each grid point. Hence the wavelet method, unlike the multitaper, should use the appropriate ‘window’ (i.e., scale) for the coherence transition wavelength at each point.

In this contribution we respond to criticism of the fan wavelet approach made by Pérez-Gussinyé et al. (2007, 2009a) that its lower spatial resolution than the multitaper method led to $T_e$ features being over-smoothed. Both Pérez-Gussinyé et al. (2007) and Pérez-Gussinyé et al. (2009a) presented maps of South American $T_e$ using the multitaper method which revealed short-wavelength $T_e$ variations not resolved in
the fan wavelet-derived map of Tassara et al. (2007). Hence, we address this acknowledged shortcoming by employing a Morlet wavelet of higher spatial resolution in the fan wavelet transform. We also follow Kirby and Swain (2009) and invert the square of the real part of the wavelet coherency, rather than the coherence, because it is less sensitive to correlations between the initial loads on the plate, and to “gravitational noise”, both of which can cause incorrect recovery of $T_c$ (Kirby and Swain, 2009).

To demonstrate the improvement gained in using the complete Morlet wavelet, we perform resolution and accuracy tests of the new method on synthetic data, and compare these with results from application of the multitaper method.

2. The wavelet transform

2.1 The fan wavelet transform

The wavelet transform is a space-domain convolution of a signal with a wavelet, at a series of scales. In the continuous wavelet transform (CWT), the scales are chosen arbitrarily, but here we choose them to span a dyadic grid from the Nyquist wavelength to the maximum dimension of the study area (e.g., Kirby, 2005). The CWT is more readily computed in the wavenumber domain with the fast Fourier transform. For a 2D space-domain ($x$) signal $g(x)$ with Fourier transform $G(k)$, its wavelet coefficients are computed via:

$$\tilde{g}(s,x,\theta) = F^{-1}[G(k)\hat{\psi}_{s\theta}(k)]$$

where $k$ is 2D wavenumber, $s$ and $\theta$ are the scale and azimuth of the Morlet wavelet respectively, $\hat{\psi}_{s\theta}(k)$ is the Fourier transform of the wavelet at a certain scale and azimuth, and $F^{-1}$ is the inverse 2D Fourier transform. The wavelet coefficients are...
hence functions of position, scale and azimuth, with scale then converted to an equivalent Fourier wavenumber (see below). The wavelet coherency between two signals $g$ and $h$ is then computed from:

$$\Gamma(s, x) = \frac{\langle \tilde{g} \tilde{h}^* \rangle_\theta}{\langle \tilde{g} \tilde{g}^* \rangle_\theta^{1/2} \langle \tilde{h} \tilde{h}^* \rangle_\theta^{1/2}}$$

(3)

which is complex-valued (Kirby and Swain, 2009). The averaging in Eq. (3) is performed over azimuth, at a particular scale, which gives isotropic coefficients if $\theta$ varies from 0° to 180° (Kirby, 2005). The coherence, Eq. (1), is thus the modulus-squared of the coherency, but as mentioned in the Introduction, we estimate $T_e$ by inverting the squared-real-coherency (“SRC”), i.e., $(\text{Re } \Gamma)^2$.

The fan wavelet transform of Kirby (2005) was developed in order to give wavelet coefficients that are both isotropic and complex. Isotropic coefficients are required to avoid an anisotropic bias on the estimate of $T_e$, while the coefficients must be complex-valued to preserve phase information. In contrast, other complex-valued 2D wavelets such as the Morlet wavelet give anisotropic wavelet coefficients, while isotropic wavelets such as the derivative of Gaussian family give real-valued wavelet coefficients. The fan wavelet transform is hence a series of rotated 2D Morlet wavelet coefficients spanning 180° that are then superposed for isotropy, while preserving their complex nature.

2.2 The complete Morlet wavelet

2D Morlet wavelets have the wavenumber domain equation:

$$\hat{\psi}(k) = e^{-|k|_{/2}}$$

(4)
where \( \mathbf{k}_0 = (|\mathbf{k}_0| \cos \theta, |\mathbf{k}_0| \sin \theta) \), and \( \theta \) is the resolving direction of the wavelet (e.g., Antoine et al., 2004). While the choice of the value of the central wavenumber, \( |\mathbf{k}_0| \), of the Morlet wavelet is somewhat arbitrary, the value we have used in our previous work (\( \pi \sqrt{2/\ln 2} \approx 5.336 \)) is standard throughout much of the wavelet literature, and gives a wavelet where the amplitude of the first sidelobes of the real part of the space-domain wavelet are a half that of the amplitude of the central peak (Addison, 2002) (see top-left panel of Fig. 1). If not actually using this specific value, all authors stipulate that \( |\mathbf{k}_0| \geq 5 \) in order for Eq. (4) to represent a valid wavelet. When \( |\mathbf{k}_0| < 5 \), the mean value of the wavelet in the space domain no longer approximates zero and thus does not satisfy the zero-mean requirement of wavelets (e.g., Antoine et al., 2004). This is evident from Eq. (4) for \( |\mathbf{k}_0| = 3 \), for example: at the zero wavenumber (\( |\mathbf{k}| = 0 \)) the mother wavelet transform has a value of 0.01, a significant ‘DC component’. In contrast, when \( |\mathbf{k}_0| = 5.336 \), its value here is \(<10^{-6}\).

[FIG. 1]

However, allowing \( |\mathbf{k}_0| \) to vary can be very useful for some applications. According to Addison (2002) and Addison et al. (2002), when \( |\mathbf{k}_0| \) is relatively low, the Morlet wavelet gives a better resolution in the space-domain (“\( \mathbf{x} \)-resolution”) than for mid-range values (\( |\mathbf{k}_0| \sim 5 \)), but therefore must give a worse resolution in the wavenumber domain (“\( \mathbf{k} \)-resolution”) from the uncertainty relation:

\[
\Delta \mathbf{x} \Delta \mathbf{k} \geq 2\pi \tag{5}
\]

This improved \( \mathbf{x} \)-resolution arises because, for a desired wavenumber (e.g., the coherence/SRC transition wavenumber), a \( |\mathbf{k}_0| = 3.773 \) wavelet is of a smaller scale (smaller spatial extent) than the corresponding \( |\mathbf{k}_0| = 5.336 \) wavelet needed to resolve
this particular wavenumber. This is shown in Fig. 1, where the wavelet scale has been
chosen so that the peak wavenumber (which gives the equivalent Fourier
wavenumber) is located at the Bouguer coherence transition wavenumber. When $|k_0|$ has the lower value, the ‘transition wavelet’ (i.e., that wavelet whose equivalent
Fourier wavenumber (scale) corresponds to the coherence/SRC transition
wavenumber) has a smaller scale (giving it a better $x$-resolution), but a
commensurately larger bandwidth (giving it a worse $k$-resolution). Conversely, when
$|k_0|$ is relatively high, the resulting wavelet coefficients have a poor $x$-resolution but a
good $k$-resolution. The larger the value of $|k_0|$, the more the global scalogram tends
towards the periodogram estimate, though of course the spatial resolution decreases.

If a $|k_0| < 5$ wavelet is desired, then the ‘complete Morlet wavelet’ must be used. In
the space domain the mother wavelet equation is:

$$\psi(x) = \left( e^{k_0 x} - e^{-k_0 x^2 / 2} \right) e^{-k_0 x^2 / 2}$$  \hspace{1cm} (6)

with Fourier transform:

$$\hat{\psi}(k) = e^{-k_0 k^2 / 2} - e^{-(k_0 - k)^2 / 2}$$  \hspace{1cm} (7)

(Addison et al., 2002). The extra $e^{-|k_0|^2 / 2}$ terms in these equations correct for the
non-zero mean of the wavelet at small $|k_0|$. In general, if the amplitude of the first
sidelobes is a fraction $1/p$ (where $p > 1$) of the amplitude of the central peak of the
space domain wavelet, then $|k_0| = \pi \sqrt{2/\ln p}$. The values of $|k_0|$ that we use in this
study are 2.668, 3.081, 3.773, 5.336 and 7.547, which give a space domain wavelet
whose first sidelobes are $1/16$, $1/8$, $1/4$, $1/2$, and $1/\sqrt{2}$ of the magnitude of the central
amplitude, respectively.
When the complete Morlet wavelet is to be used in a fan geometry, as in this study, the formula for the equivalent Fourier wavenumber ($\kappa$) at a certain wavelet scale ($s$) becomes the solution to the equation:

$$\kappa \left(1 - e^{-|k_0|^2} \right) - \frac{|k_0|}{s} = 0$$  \hspace{1cm} (8)

which may be solved for $\kappa$ using the Newton-Raphson method. However, when $|k_0| = 2.668$ (the smallest value we use here), the error in $\kappa$ that is committed by using the approximation $\kappa = |k_0|/s$ (Kirby, 2005) amounts to only 0.08%. A similar consequence applies to the angular separation between the Morlet wavelets, so we restrict our analyses to values of $|k_0| \geq 2.668$. We have also confirmed that there is only a very minor bias to the wavelet spectra of fractal surfaces when using $|k_0| = 2.668$ compared to spectra from $|k_0| = 5.336$ (e.g., Kirby, 2005).

Finally, it is useful to derive an expression for the width of the wavelet in the space-domain, or “footprint”. Since the Gaussian envelope never decays to zero, we define the footprint as space-domain width at half the wavelet’s maximum amplitude. The Gaussian envelope of the daughter wavelet has the form $\psi_s(x) = \exp \left(-x^2/2s^2\right)$ from Eq. (6), so at half-maximum amplitude its width ($2x$) is:

$$\Delta_x = 2s\sqrt{2\ln 2}$$  \hspace{1cm} (9)

Relating scale to an equivalent Fourier wavelength ($\lambda_e = 2\pi/\kappa = 2\pi s/|k_0|$) we get:

$$\Delta_x = \frac{\lambda_e |k_0|\sqrt{2\ln 2}}{\pi}$$  \hspace{1cm} (10)

Eq. (10) hence gives the space-domain width of the wavelet needed to resolve an anomaly with wavelength $\lambda_e$. Remember though that the wavelet amplitude never
decays to zero and that it is convolved with all data in the study area; Eq. (10) merely describes its primary region of influence.

3. The multitaper method

The multitaper (MT) method was introduced by Thomson (1982) in order to provide better estimates of power spectra than those obtained by, for example, the periodogram. The data are first multiplied by a set of orthogonal tapers in the space domain, the Fourier transform of the data-taper product taken for each taper, and the power spectrum determined at each taper. The final estimate of the signal’s true power spectrum is then the weighted average of the individual power spectra over all tapers.

The tapers are designed to reduce spectral estimation bias and, because they are orthogonal, combining them reduces the variance. Slepian (1978) introduced a set of orthonormal tapers, discrete prolate spheroidal sequences, whose spectra comprise a central lobe of halfwidth \( W \), with sidelobes of decreasing amplitude. The ideal spectrum is, of course, a delta function, i.e., a spike with no sidelobes. Hence, decreasing \( W \) gives an increased resolution in the wavenumber-domain because the taper spectrum tends towards a delta function, but this also results in increased leakage between harmonics because the taper becomes broader in the space domain (e.g., Simons et al., 2000).

The halfwidth of a taper is commonly expressed as the product \( NW \), where \( N \) is the number of data observations. (The halfwidth is actually \( NW\Delta x \), where \( \Delta x \) is the space-domain spacing of the gridded data.) For a given \( NW \) there are \( k = 2NW - 1 \) useful
tapers, where $2NW$ is the Shannon number, beyond which the tapers have greater
leakage. As the order of the tapers increases, the number of oscillations in the space-
domain increases, and in the wavenumber-domain the central lobe broadens and the
number and amplitude of the sidelobes increases, resulting in a poorer wavenumber-
domain resolution (e.g., Simons et al., 2000). However, using higher-order tapers is
useful because the final estimation variance of the multitaper power spectrum
decreases as $1/k$ (e.g., Simons et al., 2000, 2003; Prieto et al., 2007).

When applied to $T_e$ estimation, the 2D MT Fourier transform must be used, giving $k^2$
tapers in total, i.e., $k$ in each direction. McKenzie and Fairhead (1997) used $k = 3$
tapers, with $NW = 4$ (D. McKenzie, personal communication, 11 August 2010).
Simons et al. (2000) applied the method to several regions within Australia using $k = 7$ and $NW = 4$. Swain and Kirby (2003) also analysed Australian $T_e$, though using $k = NW = 2$. In these studies, however, $T_e$ was estimated within a single window.
Alternatively, the spatial variations in $T_e$ can be mapped by applying the multitaper
method to many windows covering the study area. Pérez-Gussinyé et al. (2004) thus
mapped $T_e$ variations over Fennoscandia using $k = 3$ and $5$, and $NW = 3$, and windows
of size $1000\times1000$ km, and $1200\times1200$ km. The moving-window MT method was
also applied to the Canadian Shield by Audet and Mareschal (2004), who used $k = NW = 3$, and windows of size $1024\times1024$ km. In more recent studies, Pérez-Gussinyé
et al. (2007) used $k = 5$ and $NW = 3$, and windows of size $400\times400$ km, $600\times600$ km,
and $800\times800$ km, while Pérez-Gussinyé et al. (2009a) used $k = NW = 3$, and windows
of size $600\times600$ km, with both studies looking at South American $T_e$. 
While small windows afford improved spatial resolution, they cannot capture large coherence transition wavelengths associated with high $T_e$ values. The resulting downward bias in $T_e$ estimates is reflected in Fig. 2 of Pérez-Gussinyé et al. (2009a). In this study we apply the MT method to synthetic data using $k = NW = 3$, and a window of size $400 \times 400$ km that is moved by the data grid spacing (i.e., every 20 km). This relatively small window size is chosen to test the spatial resolution limits of the MT method, when compared against the new complete Morlet wavelet method. We show below (Section 6.1) that a simple modification to the predicted coherence formula considerably reduces the resulting small-window bias.

4. Inversion for $T_e$

4.1 Wavelet transform

$T_e$ is recovered from the Bouguer anomaly and topography using the wavelet transform method in the following manner. First, both datasets are mirrored about their edges prior to Fourier transformation, which, when used with the wavelet transform does not generally bias the results significantly, as it can with the periodogram method (Kirby and Swain, 2008). An observed SRC is then calculated from these data, then inverted using the “load-deconvolution” method of Forsyth (1985), whereby the initial loads are recreated assuming a starting $T_e$ value in the thin elastic plate equations, and a predicted SRC determined (see the Appendix, and Swain and Kirby, 2006). In these synthetic experiments, we use the same crustal/mantle densities and depths as used in the synthetic model generation (Section 5). The minimum $\chi^2$ misfit between observed and predicted SRC is found using Brent’s method of minimisation (Press et al., 1992), with the inversion weighted using jackknife error estimates (Thomson and Chave, 1991). The $T_e$ value corresponding to
the minimum $\chi^2$ misfit is assigned to the grid node, and the procedure repeated for all
grid nodes in the study area. The $\chi^2$ misfit curves for a range of $T_e$ values also yield an
estimate of the error on $T_e$, represented by, for example, the 95% confidence limits
(Press et al., 1992; Kirby and Swain, 2009). However, for visual clarity in the figures
we do not show these limits here.

Note that the load-deconvolution method can also estimate the ratio between the
initial internal and surface load amplitudes (the “loading ratio”, $f$). Since this study
focuses on $T_e$ recovery we do not discuss estimation of $f$ here, though the interested
reader can find this in Kirby and Swain (2008, 2009).

4.2 Multitaper method

Inversion of the observed SRC for $T_e$ using the MT method is approached in a slightly
different manner, in that the data area is segmented into many overlapping windows,
and the load deconvolution applied within each window. That is, the gravity and
topography data are windowed, Fourier transformed, and the initial loads determined
in the wavenumber domain from a starting $T_e$ value. The MT and wavelet methods
then depart in what happens next. In the MT approach, the initial load transforms
must be inverse Fourier transformed back to the space domain, in order to be
multiplied by the tapers, and then Fourier transformed back to the wavenumber
domain, in order to compute the predicted SRC for the window (see the Appendix).
These extra steps are necessary so that the same procedure that was applied to the data
in order to compute an observed SRC is also applied to the predicted initial loads to
compute a predicted SRC. If these measures are not taken, Pérez-Gussinyé et al.
(2004) found that biases between observed and predicted spectra arise, which then bias $T_e$ estimates.

As for the wavelet approach, we find the minimum $\chi^2$ misfit between observed and predicted SRC using Brent’s method of minimisation (Press et al., 1992), with the inversion weighted using the analytical formula for coherence error (e.g., Bechtel et al., 1987). Again, we use the same crustal/mantle densities and depths as used in the synthetic model generation (Section 5). The $T_e$ value corresponding to the minimum $\chi^2$ misfit is assigned to the window, and the procedure repeated for all windows covering the study area. $\chi^2$ confidence limits on $T_e$ can also be computed (Section 4.1). As for the computation of observed MT SRC (Section 3), we use $k = NW = 3$, a window of size $400 \times 400$ km, with the windows moved by 20 km in both easting and northing.

In the Appendix we present an approach by which to improve $T_e$ estimates from the multitaper method. Briefly, the new approach involves performing the load deconvolution on the observed gravity, rather than on its downward continued approximation to Moho relief. Results of this modification are presented in Section 6.1.

5. Synthetic models

The performance of any $T_e$-recovery method may be ascertained by applying it to various synthetic plate models with known $T_e$ (e.g., Macario et al., 1995; Kirby and Swain, 2008). In our method, two random, fractal surfaces are generated using the method of Peitgen and Saupe (1988), using different random seeds for each, over an
area of 5100×5100 km, on a 20 km grid. Each surface has fractal dimension 2.5, and
together act as the initial surface and internal loads upon a thin elastic plate. In all our
synthetic models, the internal load is applied at a Moho of depth 35 km, and the crust
comprises a single layer of density 2800 kg.m\(^{-3}\) overlying a mantle of density 3300
kg.m\(^{-3}\). We also always assign the loading ratio, \(f\), to be 1. The final Bouguer anomaly
and topography after loading are then obtained by solving the flexural equation of a
thin elastic plate. For plates with a uniform \(T_e\) distribution the flexural equation is
solved using the Fourier transform, while for plates with a spatially-variable \(T_e\)
distribution the method of finite differences is used (see Kirby and Swain, 2008).

This complete process is then repeated a further 99 times for the chosen \(|k_0|\) value
using a different ‘seed’ in the random number generator each time. This gives
different random, fractal load surfaces. After inversion of the 100 gravity/topography
pairs, the resulting 100 \(T_e\) estimates are then averaged, in order to reduce the effect of
random and unavoidable initial load correlations (Kirby and Swain, 2008).

6. Resolution and accuracy results from synthetic modelling

6.1 Variable-\(T_e\) plate

In order to demonstrate the increase in spatial resolution gained by using wavelets
with a lower value of \(|k_0|\), we performed synthetic modelling using a plate with a
spatially-variable \(T_e\) distribution. The model \(T_e\) is derived from a chirp signal, with
spatially decreasing wavelength (Fig. 2a).

[FIG. 2]

[FIG. 3]
Figs 2b–f show the averaged $T_e$ estimated from the SRC at different $|k_0|$ values. As expected, the lower-$|k_0|$ wavelets give a better space-domain $T_e$ resolution than the higher-$|k_0|$ wavelets, better defining the decreasing wavelength of the chirp signal. However, the $T_e$ values of all the features in Fig. 2 are underestimated with respect to the input model, shown in the cross-sections in Fig. 3.

As discussed and explained by Kirby and Swain (2008), this arises because the fan wavelet method tends to underestimate relative $T_e$ differences when the $T_e$-anomaly width is smaller than its coherence/SRC transition wavelength, and hence smaller than the footprint of the transition wavelet needed for its resolution (see Section 2.2 and Eq. (10)). Hence, when the transition wavelength is large (and $T_e$ large), the transition wavelets will be large scale and the observed SRC will be spatially smoothed. Furthermore, high-$|k_0|$ transition wavelets are spatially broader than their low-$|k_0|$ counterparts (see Fig. 1), resulting in smoother observed SRCs, and a more underestimated relative $T_e$ difference.

For a maximum $T_e = 60$ km with $f = 1$, as used here, the coherence/SRC transition wavelength is 419 km (Kirby and Swain, 2008). From Eq. (10) the transition wavelet footprint is 781 km at $|k_0| = 7.547$, but only 276 km at $|k_0| = 2.668$, explaining the decrease in $x$-resolution for the higher-$|k_0|$ wavelets as the chirp wavelength decreases (Fig. 3). The longest-wavelength chirp oscillation in the model is well-resolved by all $|k_0|$ wavelets because its wavelength is ~1500 km, i.e., greater than their footprints and not too small to be averaged out. However, the smallest chirp wavelength is only 100 km, so here even the $|k_0| = 2.668$ has trouble resolving it.
In the MT results the recovered resolution of the chirp oscillations is approximately comparable to the $|k_0| = 3.081$ or 3.773 results, though the amplitude is much reduced and biased downwards by $\sim 12$ km uniformly across the area (Figs 4a and 5). The resolution difference between the multitaper and $|k_0| = 2.668$ wavelet results at short chirp wavelengths (Fig. 5) can be explained by the larger multitaper window (400 km, compared to the wavelet footprint of 276 km) averaging over a greater area.

Of particular note is the correction of the downward $T_e$ bias in the conventional approach (see the Appendix). As can be seen in Fig. 5, the conventional approach gives $T_e$ values that have a $\sim 12$ km downward bias (green line in Fig. 5). However, by performing the load deconvolution using the observed gravity, rather than its downward continued expression, we recover the grid shown in Fig. 4b, shown as cross-section in Fig. 5 (red line). While not improving upon the spatial resolution, the new approach largely corrects for the 12 km underestimation of $T_e$. It gives a very similar correction to one based on Fig. 2a of Pérez-Gussinyé et al. (2009a,b) and assuming a uniform $T_e$ of 35 km, which is its true average value.

### 6.2 Uniform-$T_e$ plates

Synthetic modelling of a uniform-$T_e$ plate eliminates the effect discussed in the previous section (i.e., large wavelet footprint vs narrow $T_e$ anomalies) to give an unbiased accuracy assessment. We chose uniform-$T_e$ plates with values ranging from
10 km to 150 km in intervals of 10 km. We also used five different \(|k_0|\) values for the wavelet method, but did not perform inversion using the MT method, since this has been done elsewhere (e.g., Pérez-Gussinyé et al., 2009a,b).

[FIG. 6]

Fig. 6 shows the results. As expected, the \(|k_0| = 7.547\) wavelet gives the most accurate results, because on a uniform plate with no \(T_e\) anomalies its poor \(x\)-resolution is irrelevant, while its good \(k\)-resolution is better able to resolve the transition wavelength than can wavelets with smaller \(|k_0|\) values. Conversely, the \(|k_0| = 2.668\) wavelet gives the worst \(T_e\) recovery, underestimating \(T_e\) by up to 23%. Note that the percentage underestimation remains relatively constant above model \(T_e\) values of 30 km.

It must be stressed, however, that the “calibration curves” in Fig. 6 represent the bias for a large (5100×5100 km) plate of uniform \(T_e\). If \(T_e\) anomalies are narrow, or the study area smaller, then the downward bias to \(T_e\) will be more pronounced (Kirby and Swain, 2008). It is for this reason that we do not apply a correction to our \(T_e\) estimates based upon these calibration curves, as did Pérez-Gussinyé et al. (2009a,b) for the multitaper method.

Nevertheless, we note that even the \(|k_0| = 2.668\) curves in Fig. 6a give much less bias than the 400×400 km windows of the multitaper method, as shown in Fig. 2a of Pérez-Gussinyé et al. (2009b). We recognise, however, that implementation of our
new multitaper load deconvolution approach should improve upon the $T_e$ estimates of Pérez-Gussinyé et al. (2009b).

7. South American $T_e$

To provide a real data example of the new method, Fig. 7 shows $T_e$ computed over South America using $|k_0| = 2.668$ and 5.336 wavelets. The 5.336 result is almost identical to Fig. 3a of Tassara et al. (2007) (allowing for the different colour scales). Noticeable differences between the 2.668 and 5.336 results occur in the Guyana shield north of the Tacutu rift, and along the Transbrasiliano lineament.

[FIG. 7]

A criticism of the ($|k_0| = 5.336$) wavelet approach made by Pérez-Gussinyé et al. (2007) was that its lower resolution than their multitaper results led to $T_e$ features in South America being smoothed out. The examples they cited were the low $T_e$ along a substantial part of the Transbrasiliano lineament, the dyke swarms associated with the Paraná flood basalts, the Amazon basin, and the Tacutu rift. This criticism is, of course, valid for the 5.336 results, but is no longer so for the 2.668 results. A comparison of our Fig. 7a with Fig. 6a of Pérez-Gussinyé et al. (2009a) shows that the new wavelet method has succeeded in resolving very similar features in all these areas.

Although it is outside the scope of this paper, it should be noted that gravitational “noise” (McKenzie and Fairhead, 1997; McKenzie, 2003; Kirby and Swain, 2009) does exist over the continent, and casts doubt upon $T_e$ values in some regions,
especially where we have recovered very high values. We return to this issue in the
Conclusions.

8. Conclusions

We have shown how the fan wavelet method for $T_c$ estimation may be improved to
attain increased spatial resolution. Although this comes at the expense of a decreased
wavenumber resolution, our synthetic modelling results show that this does not lead to
significantly poorer $T_c$ estimates. This refutes the arguments of Pérez-Gussinyé et al.
(2007, 2009a) that the fan wavelet method inherently had poorer spatial resolution
than the multitaper approach. Indeed, the wavelet method has significant advantages
over the multitaper method. First, the fan wavelet method only requires the arbitrary
variation of one parameter ($|k_0|$), whereas the multitaper method requires the variation
of three (window size, bandwidth and number of tapers). The fan wavelet method is
also an order of magnitude computationally faster than the multitaper method
because, as described in Section 4.2 and the Appendix, there are fewer arrays to store
and fewer Fourier transforms to compute.

Second, the wavelet method allows for a more accurate estimation of $T_c$ when using
theoretical formulae. As Pérez-Gussinyé et al. (2004) demonstrated, observed
cohereces and admittances are increasingly biased as the window size decreases, due
to spectral leakage, yet small windows are needed to resolve smaller $T_c$ anomalies.
This restriction presents no problems if a load-deconvolution approach is adopted,
where predicted coherences/admittances are computed using windows of the same
size, so that the leakage is the same for both observed and predicted quantities.
However, theoretical coherence/admittance formulae are derived via the continuous
Fourier transform, which assumes an infinite data area without spectral leakage. So when observed, windowed coherences/admittances are compared with theoretical predictions, the resulting best-fit $T_e$ will be biased. This is not the case for the wavelet transform if a large, (e.g., continent-sized) study area is chosen, because the wavelet method does not explicitly window the data. The resulting coherences/admittances have minimal bias when compared with theoretical coherence/admittance formulae (Kirby and Swain, 2004).

We have also reduced the downward bias of $T_e$ estimates from the multitaper method, by performing the load deconvolution on gravity rather than Moho topography. This, however, does not improve the spatial resolution of the MT method, which remains slightly poorer than that of the wavelet transform approach with complete Morlet wavelets of low central wavenumber.

Finally, we provide some recommendations for $|k_0|$ values. Owing to the better $k$-resolution, values of $|k_0| > 5$ give more accurate absolute $T_e$ estimates if the tectonic province is large and contains a relatively uniform $T_e$; in such cases, the low $x$-resolution of these wavelets is of less importance. If detailed $T_e$ structure and/or accurate relative $T_e$ differences are required, then values of $|k_0| < 3.5$ are more likely to be useful, though the lower $k$-resolution of these wavelets could indicate greater uncertainty on the $T_e$ estimates.

There remains, however, the issue of “gravitational noise” which can contaminate the Bouguer coherence and render load-deconvolution methods invalid (McKenzie, 2003; Kirby and Swain, 2009). This problem can occur in areas of subdued topography.
Although we do not show it here, we have found: 1) that such noise predominantly affects regions of the Earth where the coherence method indicates high $T_e$ (Kirby and Swain, 2009); 2) that the size of the area affected increases as $|k_0|$ decreases; and 3) that the lower $k$-resolution of low-$|k_0|$ wavelets results in the noise spectrum being smeared over a larger bandwidth than with high-$|k_0|$ wavelets. Hence, although low-$|k_0|$ wavelets afford better spatial resolution, they are most useful when applied to the younger regions of the continents where $T_e$ is more likely to be low.

The $T_e$ estimation software is available at no cost as Fortran95 source code from the corresponding author.

Appendix. Load deconvolution

In the load deconvolution method of Forsyth (1985), initial surface and internal (i.e., within or at the base of the crust) loading processes may be considered separately, then superposed. The following equations are obtained in the Fourier domain.

Consider two initial loads of geometric amplitude $H_i$ and $W_i$ applied at the surface and Moho (at depth $z_m$ from sea level), respectively, of a thin elastic plate of known $T_e$.

The initial surface load produces new surface topography $H_T$, and new Moho topography $W_T$. Similarly, the initial internal load produces new surface topography $H_B$, and new Moho topography $W_B$. The relationships between the final surface/Moho topographies and the initial loads are:

$$W_B = v_B W_i$$
$$W_T = v_T H_i$$
$$H_B = \kappa_B W_i$$
$$H_T = \kappa_T H_i$$  \(\text{(A1)}\)
where the wavenumber-dependent deconvolution coefficients are obtained from the solution to the thin elastic plate equations (Forsyth, 1985):

\[ v_B = 1 - \frac{\Delta \rho_1}{\Phi} \]
\[ v_T = -\frac{\Delta \rho_1}{\Phi} \]
\[ \kappa_B = -\frac{\Delta \rho_2}{\Phi} \]
\[ \kappa_T = 1 - \frac{\Delta \rho_1}{\Phi} \]  
(A2)

where \( \Delta \rho_1 = \rho_c - \rho_f \), \( \Delta \rho_2 = \rho_m - \rho_c \), \( \rho_c \) and \( \rho_m \) are crust (2800 kg.m\(^{-3}\)) and mantle (3300 kg.m\(^{-3}\)) densities, respectively, \( \rho_f \) is the density of the overlying fluid (air or water), and where:

\[ \Phi = \frac{D|k|^4}{g} + \rho_m - \rho_f \]  
(A3)

\( D \) is flexural rigidity, |\( k \)| is the 1D wavenumber, \( g \) is the gravitational acceleration (9.8 ms\(^{-2}\)). The flexural rigidity (D) is related to the elastic thickness (\( T_e \)) by:

\[ D = \frac{E T_e^3}{12(1-\sigma^2)} \]  
(A4)

with \( E \) being Young’s modulus (100 GPa), and \( \sigma \) being Poisson’s ratio (0.25).

The total resultant surface topography, \( H \), if both processes act together will be \( H = H_T + H_B \), while the net Moho topography \( W = W_T + W_B \). For combined surface and internal loading, the initial loads are related to the final, predicted Moho and surface topographies after flexure (\( W \) and \( H \), respectively) by a matrix equation:

\[ \begin{pmatrix} W \\ H \end{pmatrix} = \begin{pmatrix} v_B & v_T \\ \kappa_B & \kappa_T \end{pmatrix} \begin{pmatrix} W_i \\ H_i \end{pmatrix} \]  
(A5)

[c.f. Eq. (18) of Forsyth (1985)].
In Forsyth (1985)’s original method, the observed Moho topography \((W)\) is determined from the observed Bouguer gravity anomaly \((G)\) using the relationship:

\[
W = \frac{G e^{H_s}}{2\pi G \Delta \rho_2} \tag{A6}
\]

where \(G\) is the Newtonian gravitational constant (Parker, 1972). A predicted coherence (for an assumed \(T_e\)) is then determined by substitution of \(H = H_T + H_B, W = W_T + W_B\) and Eq. (A6) into the coherence formula, Eq. (1), though here we use the predicted coherency [c.f., Eq. (3)]:

\[
\Gamma_p^{(W)}(\|k\|) = \frac{\langle W_T H_T^* + W_B H_B^* \rangle_{\|k\|}}{\langle W_T W_T^* + W_B W_B^* \rangle_{\|k\|}^{1/2} \langle H_T H_T^* + H_B H_B^* \rangle_{\|k\|}^{1/2}} \tag{A7}
\]

where \(\langle \cdot \rangle_{\|k\|}\) indicates averaging over annuli of equal wavenumber (though see discussion below), and Forsyth (1985) assumed that \(\langle \exp(-kz_m) W(k) \rangle_{\|k\|} = \exp(-|k|z_m) \langle W(k) \rangle_{\|k\|}\), so that the exponential terms cancel. Forsyth (1985) also assumed that the surface and internal loading processes are independent, or statistically uncorrelated. This restriction is expressed by setting the average of terms containing both surface \((T)\) and internal \((B)\) loads to zero, e.g., \(\langle W_T H_B^* \rangle = 0\), etc.

Eq. (A7) is the formula used by the MT method for predicted coherency computation, though the averaging is also performed over the tapers, and then over wavenumber annuli.
An alternative approach is to compute the surface and internal loading components of the Bouguer anomaly, rather than Moho topography. In this fashion, we have:

\[ G_B = \mu_B W_i \]
\[ G_T = \mu_T H_i \]
\[ H_B = \kappa_B W_i \]
\[ H_T = \kappa_T H_i \]  \hfill (A8)

where the wavenumber-dependent gravity deconvolution coefficients are:

\[ \mu_B = \frac{2\pi G \Delta \rho_z}{\Phi} \left( \Phi - \Delta \rho_z \right) e^{-\|k\| \Phi} = 2\pi G \Delta \rho_z e^{-\|k\| \Phi} \]
\[ \mu_T = \frac{-2\pi G \Delta \rho_z}{\Phi} \Delta \rho_z e^{-\|k\| \Phi} = 2\pi G \Delta \rho_z e^{-\|k\| \Phi} \]  \hfill (A9)

and the expressions for \( \kappa_B \) and \( \kappa_T \) in Eq. (A2) do not change. The total resultant surface topography, \( H \), if both processes act together will be \( H = H_T + H_B \), while the net Bouguer anomaly is \( G = G_T + G_B \). For combined surface and internal loading, the initial loads are related to the final, predicted Bouguer anomaly and surface topography after flexure by:

\[ \begin{pmatrix} G \\ H \end{pmatrix} = \begin{pmatrix} \mu_B & \mu_T \\ \kappa_B & \kappa_T \end{pmatrix} \begin{pmatrix} W_i \\ H_i \end{pmatrix} \]  \hfill (A10)

The predicted coherency is then:

\[ \Gamma_p^{(G)} (|k|) = \frac{\left\langle G_T H_T^* + G_B H_B^* \right\rangle_{|k|}}{\left\langle G_T G_T^* + G_B G_B^* \right\rangle_{|k|}^{1/2} \left\langle H_T H_T^* + H_B H_B^* \right\rangle_{|k|}^{1/2}} \]  \hfill (A11)

In theory, Eqs (A7) and (A11) should give identical predicted coherencies. However, in practice they do not (as shown in Figs 4 and 5). If Eqs (A1), (A5) and (A6) are substituted into Eq. (A7) we obtain the predicted coherency as computed by deconvolution using Moho topography:
\[\Gamma_p^{(W)}(|k|) = \frac{\langle e^{\mathbf{j} \mathbf{H}_0 \Delta_w^{-2} (v_t \kappa_T \eta \eta^* + v_b \kappa_B \zeta \zeta^*)} \rangle_{|k|}}{\langle e^{\mathbf{j} \mathbf{H}_0 \Delta_w^{-2} (v_t^2 \eta \eta^* + v_b^2 \zeta \zeta^*)} \rangle_{|k|}^{1/2} \langle e^{\mathbf{j} \mathbf{H}_0 \Delta_w^{-2} (\kappa_T^2 \eta \eta^* + \kappa_B^2 \zeta \zeta^*)} \rangle_{|k|}^{1/2}} \quad (A12)\]

where \(\Delta_w = v_b \kappa_T - v_t \kappa_B\) is the determinant from Eq. (A5); \(\eta = -\kappa_B G + A e^{\mathbf{j} \mathbf{H}_0} v_b H\);

and \(\zeta = \kappa_T G - A e^{\mathbf{j} \mathbf{H}_0} v_t H\). The expression when the gravity-deconvolution method is used is obtained by substituting Eqs (A8) and (A10) into Eq. (A11), and rearranging to match Eq. (A12) as best as possible:

\[\Gamma_p^{(G)}(|k|) = \frac{\langle e^{\mathbf{j} \mathbf{H}_0 \Delta_w^{-2} (v_t \kappa_T \eta \eta^* + v_b \kappa_B \zeta \zeta^*)} \rangle_{|k|}}{\langle \Delta_w^{-2} (v_t^2 \eta \eta^* + v_b^2 \zeta \zeta^*) \rangle_{|k|}^{1/2} \langle e^{\mathbf{j} \mathbf{H}_0 \Delta_w^{-2} (\kappa_T^2 \eta \eta^* + \kappa_B^2 \zeta \zeta^*)} \rangle_{|k|}^{1/2}} \quad (A13)\]

Comparison of Eqs (A12) and (A13) shows that the former has extra terms of \(\exp(|k|z_m)\) and \(\exp(2|k|z_m)\). If the annuli were narrow enough to assume but one value of \(|k|\) then these terms could be taken out of the averaging brackets and would cancel, making Eqs (A12) and (A13) identical. However, this does not occur in most practical implementations, and the exponential terms become prohibitively large, biasing the predicted coherency and hence recovered \(T_e\).

The fan wavelet method is based on a further development of Eq. (A11) which greatly improves computation time. If Eqs (A8) are substituted into Eq. (A11), we get:

\[\Gamma_p^{(G)}(|k|) = \frac{\langle \mu_T \kappa_T |H|^2 + \mu_B \kappa_B |W|^2 \rangle_{|k|}}{\langle \mu_T^2 |H|^2 + \mu_B^2 |W|^2 \rangle_{|k|}^{1/2} \langle \kappa_T^2 |H|^2 + \kappa_B^2 |W|^2 \rangle_{|k|}^{1/2}} \quad (A14)\]

as the Fourier transform expression. But with the wavelet transform the deconvolution coefficients can be taken outside the averaging brackets, because the averaging is performed over the wavelet azimuth and these coefficients are functions of wavelet
scale (equivalent Fourier wavenumber) only. So, in terms of the wavelet transform, Eq. (A14) becomes:

\[
\Gamma_p^{(07)}(s, x) = \frac{\mu_T \kappa_T \left( \langle \tilde{h}_i^2 \rangle_\theta \right) + \mu_B \kappa_B \left( \langle \tilde{w}_i^2 \rangle_\theta \right)}{\left[ \mu_T^2 \left( \langle \tilde{h}_i^2 \rangle_\theta \right) + \mu_B^2 \left( \langle \tilde{w}_i^2 \rangle_\theta \right) \right]^{1/2} \left[ \kappa_T^2 \left( \langle \tilde{h}_i^2 \rangle_\theta \right) + \kappa_B^2 \left( \langle \tilde{w}_i^2 \rangle_\theta \right) \right]^{1/2}} \tag{A15}
\]

for \( \tilde{h}_i(s, x, \theta) \) and \( \tilde{w}_i(s, x, \theta) \). Now we can make use of the loading ratio, \( f \), introduced by Forsyth (1985), which, for the wavelet transform, is computed through:

\[
f^2(s, x) = \frac{\left( \langle \tilde{w}_i^2 \rangle_\theta \right)}{\left( \langle \tilde{h}_i^2 \rangle_\theta \right)} \tag{A16}
\]

where we define \( r = \Delta \rho_1 / \Delta \rho_2 \). So dividing numerator and denominator of Eq. (A15) by \( \left( \langle \tilde{h}_i^2 \rangle_\theta \right) \) we find:

\[
\Gamma_p(s, x) = \frac{\mu_T \kappa_T + \mu_B \kappa_B f^2 r^2}{\left[ \mu_T^2 + \mu_B^2 f^2 r^2 \right]^{1/2} \left[ \kappa_T^2 + \kappa_B^2 f^2 r^2 \right]^{1/2}} \tag{A17}
\]

This demonstrates why the wavelet method is much faster than the multitaper method.

In the wavelet method, only \( \tilde{h}_i(s, x, \theta) \) and \( \tilde{w}_i(s, x, \theta) \) need be computed from the observed gravity and topography data at a given \( T_e \), and averaging is performed over azimuth only. In the multitaper method, however, four grids \([H_T(k), H_B(k), G_T(k), G_B(k)]\) must be computed at a given \( T_e \), then averaged over both wavenumber annuli and the multitapers.

\textbf{Acknowledgements.} We thank Martin Schimmel for useful discussion, Marta Pérez-Gussinyé for the South American province boundaries, and the two anonymous reviewers for their constructive suggestions. The figures were plotted using GMT.
(Wessel and Smith, 1998). This work was supported by ARC grant DP0878453, and is TIGeR publication number 232.

References


Kirby, J. F., 2005. Which wavelet best reproduces the Fourier power spectrum?

Computers and Geosciences 31(7), 846-864.


Fig. 1. Cross-sections through 2-D complete Morlet wavelets (red lines) in the space domain (real part, left-hand panels), and wavenumber domain (right-hand panels). Top row shows the wavelet at $|k_0| = 5.336$ and a scale of 213 km; bottom row shows the wavelet at $|k_0| = 3.773$ and a scale of 150 km. The black line in the wavenumber domain plots is the coherence/SRC for $T_e = 20$ km and $f = 1$, scaled to the maximum amplitude of the wavenumber domain wavelet for visual clarity. Hence, these wavelets are transition wavelets.
Fig. 2. $T_e$ averaged over 100 $T_e$ estimates recovered using the wavelet transform from 100 synthetic load pairs emplaced on a plate with the $T_e$ distribution shown in (a), which is a chirp signal with peak-to-peak $T_e$ amplitude of 50 km, and mean value 35 km. The following $|k_0|$ values were used: (b) 2.668, (c) 3.081, (d) 3.773, (e) 5.336, and (f) 7.547. $T_e$ contours in (b)-(f) at 5 km intervals.
Fig. 3. Average $T_e$ from west to east of the input $T_e$ distribution in Fig. 2a (black), and the wavelet $T_e$ grids in Figs 2b-f, for the indicated $|k_0|$ values. That is, values of constant easting are averaged.
Fig. 4. $T_c$ averaged over 100 $T_c$ estimates recovered using the multitaper Fourier transform from 100 synthetic load pairs emplaced on a plate with the $T_c$ distribution shown in Fig. 2a. Results are from deconvolution of (a) Moho relief, and (b) Bouguer anomaly – see Appendix. Multitaper window size is 400×400 km, incremented every 20 km, with $k = NW = 3$. $T_c$ contours at 5 km intervals.
Fig. 5. Average $T_e$ from west to east of the input $T_e$ distribution in Fig. 2a (black), and the multitaper $T_e$ grids in Fig. 4a (green, Moho relief deconvolution), and Fig. 4b (red, gravity deconvolution). The $|k_0| = 2.668$ wavelet cross-section from Fig. 3 is shown in blue for comparison.
Fig. 6. Uniform-$T_e$ recovery from the wavelet transform for the $|k_0|$ values indicated in (a). (a) Absolute recovered values vs input $T_e$ of the plate. The black line shows the ideal, expected relationship. (b) The difference between the average recovered and model values, as a percentage of the model $\left[\frac{100 \times (T_e^{\text{mod}} - T_e^{\text{rec}})}{T_e^{\text{mod}}}\right]$. Error bars are one standard deviation of the percentage difference $\left(\frac{100 \sigma[T_e^{\text{rec}}]}{T_e^{\text{mod}}}ight)$, from propagation of variances).
Fig. 7. $T_e$ (in km) over South America from wavelets with $|k_0| = (a) 2.668$, and (b) 5.336. We inverted the Bouguer coherence using the loading model of Banks et al. (2001) with the internal load placed at the interface between the upper and middle crust, as given by the CRUST2.0 model (Bassin et al., 2000). Topography shaded relief superimposed. Grey lines (taken from Pérez-Gussinyé et al., 2007) mark the locations of the following tectonic features: the Precambrian Guyana (Guy) and Guaporé (Gua) shields, which, together with the intervening Precambrian-underlain Amazon basin (Am), form the Amazonian craton; Precambrian São Francisco (SF) and Rio de la Plata (RP) cratons; the Precambrian-underlain Paraná basin (Par); Transbrasiliano lineament (TBL); Mesozoic Tacutu rift (Tac); Phanerozoic Patagonian terrane (Pat).